Comment sections and their role in a democratic society

Magnus André Knustad

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2023

UNIVERSITY OF BERGEN

Comment sections and their role in a democratic society

Magnus André Knustad

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 27.09.2023

Year: 2023

Title: Comment sections and their role in a democratic society

Name: Magnus André Knustad

Print: Skipnes Kommunikasjon / University of Bergen

Acknowledgements

This thesis would not be possible without the help of my supervisor Jill Rettberg and my co-supervisor Hallvard Moe.

I would like to thank my great colleagues at the Department of Linguistic, Literary and Aesthetic Studies at the University of Bergen, especially those at the Digital Culture program.

And finally, a special thanks to Daniel Apollon and Anders Løvlie for their thoughtful and thorough feedback at my halfway evaluation and masterclass.

Abstract

Newspaper comment sections provide readers with a public platform to voice their opinion on a wide range of topics, provide a direct line of feedback for journalists and editors, and have the potential of facilitating a democratically valuable public debate. However, comment sections have come under scrutiny for the prevalence of disinhibited behavior, uncivil and impolite comments, as well as politically polarizing content. In the public debate, comment sections are often described as problematic, and most research that relates to comment sections, tend to focus on incivility and impoliteness.

This thesis explores the role of comment sections in a democratic society. When considering comment sections through frameworks based on democratic theories, comment sections appear to fail to live up to democratically valuable standards. Comment sections tend to be judged by the standards of theories such as deliberative democracy and discursive, which emphasize open participation and places high value on making decisions based on reasonable argumentation. Using these theories, however, might be problematic. It is difficult to use these theories as a framework for discussing comment sections, because comment sections do not have a set point when a decision is made based on a preceding discussion. A discussion in a comment section only ends when all commenters have said what they wanted to say, at which point the debate dies down on its own without any decision having been made. Comment sections might be more suited within democratic frameworks that focus more on participation, such as participatory liberal theory and agonistic democracy. Participatory theories focus more on the participation aspect of democracy, and comment sections do, at least on first glance, make participation in the public debate easier. However, these theories also emphasize mutual respect as a basis for public discussion, something that comment sections are criticized for lacking. In the end, it might be that the best theory to understand the role of comment sections in a democratic society is the idea of the post-democracy, in which comment sections may serve a role as an anti-establishment, non-professional forum on professional, establishment news sites.

For this thesis, three topics of interest have been investigated in three papers: the effect of anonymity on toxicity, accusations of trolling, and media criticism in comment sections. This thesis presents these research projects and discusses the role of comment sections in a democratic society, as well as the methodological challenges when researching comment sections. As toxicity is a much-debated topic, and anonymity is often used to explain such behavior, a study was devised where anonymous and non-anonymous comments from the same platform were analyzed, showing that anonymity has a small, but statistically significant effect on toxicity. This thesis also found that accusations of trolling are often politically motivated and used to dismiss opposing arguments and that these accusations were mostly ignored by other debaters and the accused. Finally, this thesis explores and categorizes criticism of the media found in comment sections. Three kinds of media criticism were identified: criticism of focus, quality and of integrity. A second dimension, target of criticism, was also identified: journalists, news organizations, and the media.

The thesis concludes that the role of comment sections in a democratic society is challenging and that the greatest obstacle for comment sections playing an important, positive role is the prevalence of toxic disinhibition. There is, however, great potential for comment sections being a democratically valuable forum for public expression that incentivizes people to engage with the news media, where people have their opinions challenged and a platform for constructive criticism of the media.

Sammendrag

Kommentarfelt lar lesere uttrykke seg offentlig innen en rekke temaer, gjør det mulig med direkte tilbakemelding til journalister og redaktører, og de kan potensielt legge til rette for en demokratisk verdifull offentlig debatt. Til tross for dette er kommentarfelt blitt kritisert på grunn av uhemmet atferd, usiviliserte og uhøflige kommentarer, samt politisk polariserende innhold. I den offentlige debatten blir kommentarfelt ofte beskrevet som problematiske, og det meste av forskning relatert til kommentarfelt setter søkelys på slik uhemmet atferd

Denne avhandlingen utforsker rollen kommentarfelt har i et demokratisk samfunn. Når man ser på kommentarfelt gjennom rammeverk basert på demokratiske teorier kan det virke som at kommentarfelt ikke lever opp til demokratiske standarder. Kommentarfelt har en tendens til å bli dømt basert på standardene til deliberative demokratiske teorier. Slike teorier legger vekt på åpen deltakelse og verdsetter beslutningstaking basert på rimelig argumentasjon. Å benytte slike teorier kan derimot være problematisk. Det vanskelig å bruke deliberative teorier som et rammeverk fordi kommentarfelt ikke har et spesifikt punkt der en beslutning blir tatt på bakgrunn av den foregående diskusjonen. En diskusjon i et kommentarfelt tar slutt når alle deltakere har sagt det de skulle si, slik at debatten dør på egen hånd uten at noen beslutninger har blitt tatt. Et annet sett med demokratiske teorier som kanskje passer kommentarfelt bedre, som for eksempel *participatory liberal theory* og *agonistic democracy,* fokuserer mer på deltakelse som viktig for demokratier. Kommentarfelt gjør i første øyekast deltakelse i offentlige debatter enklere. Men slike teorier fokuserer også på gjensidig respekt som et grunnlag for offentlig debatt, noe kommentarfelt er kritisert for å mangle. Det kan være at den beste teorien for å forstå kommentarfelts rolle i et demokratisk samfunn er ideen om post-demokrati, der kommentarfelt kan ha en rolle som et anti-establishment, ikke-profesjonelt forum på profesjonelle nyhetsnettsteder.

I denne avhandlingen er tre interessefelt blitt forsket på gjennom tre artikler: effekten av anonymitet på antisosial atferd, anklagelser av trolling, og mediekritikk i kommentarfelt. Avhandlingen presenterer disse forskningsprosjektene og diskuterer kommentarfelts rolle I et

demokratisk samfunn, samt de metodologiske utfordringene som følger med når man forsker på kommentarfelt. Siden antisosial atferd blir diskutert mye og anonymitet ofte blir brukt for å forklare slik atferd, ble en studie gjennomført der anonyme og ikke-anonyme kommentarer fra samme plattformer ble analysert. Anonymitet hadde en liten, men statistisk signifikant effekt på antisosial atferd. Avhandlingen har også funnet at anklagelser av trolling ofte var politisk motivert og brukt for å se bort fra andres argumenter man ikke var enige i, og at disse anklagelsene stort sett ble ignorert av andre deltakere og de som ble anklaget. Til slutt utforsker og kategoriserer avhandlingen kritikk av media i kommentarfelt. Tre typer kritikker blitt identifisert: kritikk av fokus, kvalitet og integritet. En andre dimensjon, målet for kritikk, ble også identifisert: journalister, nyhetsorganisasjoner, og media.

Denne avhandlingen konkluderer med at kommentarfelts rolle i et demokratisk samfunn er utfordrende, og at det største hinderet for at kommentarfelt skal spille ne viktig positiv rolle er antisosial atferd. Men kommentarfelt har stort potensial til å kunne bli en demokratisk verdifull form for offentlige ytringer som får folk til å besøke nyhetsnettsteder, det er en plattform der folks meninger blir utfordret, og en plattform for konstruktiv kritikk av media.

List of papers

Knustad, M., Johansson, C. (2021) Anonymity and Inhibition in Newspaper Comments. *Information, 12(3):106.* https://doi.org/10.3390/info12030106

Knustad, M. (2020). Get Lost, Troll: How Accusations of Trolling in Newspaper Comment Sections Effect the debate. *First Monday, 25(8).* https://doi.org/10.5210/fm.v25i8.10270

Knustad, M. Media criticism in newspaper comment sections: Do comment sections constitute a democratically valuable forum for constructive criticism of the media?

Table of Contents

1. Introduction

This thesis explores three topics related to the role of comment sections in a democratic society: how anonymity affect the toxicity in comment sections, how accusations of trolling affect the debates in comment sections, and finally, how the media is criticized in comment sections.

In the online world we can reach a potentially vast audience with our opinions, be they thoughtful and argumentative, or reactive and hostile. Creating a channel for public expression is remarkably easy. The world wide web is filled with services that facilitate public expressions. There are countless forums dedicated to all sorts of topics, from specific niche topics to broad forums with hundreds and sometimes thousands of people participating actively. Various social media platforms allow their users to build an audience by connecting to other people, be they real-life friends or online strangers. Some social media platforms, such as Facebook, are closed, where what a person shares may be locked off to strangers. Others, such as Twitter, are open and public, allowing everyone to see each other's tweets. Internet users may also create blogs on a number of platforms or host their own blog. And it has also become increasingly easier for the average internet users to create their very own websites.

This thesis is about discourse in a specific kind of online forum: newspaper comment sections. While forums can be limited in their scope, and social media platforms, blogs and personal websites require an audience to be built over time, comment sections have several key features that distinguish them from other forms of public expression. Firstly, newspaper comment sections are attached to articles, making their topic of discussion limited (though, off-topic discussions and spam can certainly occur). This makes comment sections a very effective forum for expressing one's opinions about specific topics. Articles on politics, for example, are a tempting forum for expressing oneself concerning political issues. Secondly, it is relatively easy to join a discussion in a comment section. Some news sites use a Facebook plugin as their comment section, making participation as easy as clicking a button for anyone with a Facebook account. Other sites use their own platform, where users are required to create an account to comment. But the feature that perhaps most distinguishes comment sections from other forms of public expression is the potential audience size. It is difficult and time-consuming to build an audience on social media platforms, and the size of the audience in an online forum is something that the participants have little control over. The audience of a comment section, however, is the readers of the news article, thereby providing commenters with a potentially massive audience if they comment on articles published by major, national newspapers.

Comment sections represent an opportunity for people to debate each other in a public forum and to engage with journalists and editors by providing feedback on a newspaper's articles, all with a potentially massive audience. They are, however, often criticized for being forums of various antisocial behavior, such as trolling and bullying (Rowe 2014; Gonçalves 2015). It is therefore important to understand what role comment sections serve in a democratic society, especially in a time when comment sections are accused of being places of misbehavior and trolling – which in turn has led many news sites to permanently close their comment sections. In this thesis, I will investigate what role comment sections have in a democratic society. When reviewing existing theories of democracy, it becomes clear that comment sections do not live up to the standards set out in these theories. That does not mean, however, that comment sections cannot perform a democratically valuable service.

Therefore, this thesis seeks to explore various issues related to comment sections and democracy. My goal is to contribute to the understanding of the role of newspaper comment sections as sites for public debate in a democratic society. Therefore, the main research question for this thesis is: **what role does comment sections have in a democratic society?** The word *role* can be defined as "a function or part performed especially in a particular operation or process" (Merriam-Webster 2023). The *operation* or *process* that is interesting for this thesis is public discourse, specifically by the readers on mainstream news sites who comment on news articles. In other words, I'm interested in what function comment sections have, relating to public discourse, in a democratic society. *Public discourse*, of course, is a wide term that can include all forms of public expression, such as the work done by journalists themselves. It is therefore of interest to investigate what role comment sections have in relation to journalists and the media, as this would also relate to the functionality of comment sections regarding the public discourse performed by the media itself.

To operationalize this broad main research question, this dissertation project asks three interrelated sub research questions:

- **RQ1: Does anonymity affect the toxicity of comment sections?**
- **RQ2: How do accusations of trolling in newspaper comment sections effect the debate among commenters?**
- **RQ3: How is the media criticized in comment sections?**

These research questions and the articles written to answer them are described and summarized in the following table (Table 1).

Table 1: Summary of the research project and the three articles related to it.

Study purpose	To analyze the role of newspaper comment sections as a site for public debate in a democratic society.		
Data	Comments written in 2018-19 on American mainstream news sites were used for a case study of comment sections to investigate their role in a democratic society.		
	Article 1	**Article 2**	**Article 3**
Title	Anonymity and toxicity in newspaper comments	Get Lost Troll: How Accusations of Trolling in Newspaper Comment Sections Affect the Debate	Critique of the media in newspaper comment sections
Topic	Characteristics of debates in comment sections.		Discourse about mainstream media in comment sections.
Research question	Does anonymity affect the toxicity of comment sections?	How do accusations of trolling in newspaper comment sections effect the debate among commenters?	How is the media criticized in comment sections?
Sample	1400 comments from the Washington Post and the New York Times	3851 comments from the Washington Post, the New York Times, and Politico	3851 comments from the Washington Post, the New York Times, and Politico
Method	Content analysis	Qualitative text analysis	Thematic analysis
Findings	There is a weak but statistically significant relationship between anonymity and toxic comments, but anonymity cannot fully explain toxic disinhibition in comment sections.	Accusations of trolling were used to delegitimize opposing arguments based on political views rather than on rhetorical style or behavior. Accusations were generally ignored, suggesting they are not taken seriously by most commenters.	Found three kinds of media criticism: criticism of focus (the journalist is not focusing on the right issues), of quality (the journalist did not do a thorough job) and of integrity (the journalist shows political bias or is untruthful).

In the synopsis of this thesis, I will first review existing literature that relates to comment sections, participatory journalism, toxic disinhibition, trolling, how computer-mediated communication effects people, anonymity, and finally different democratic theories. I will then write about my experiences of researching comment sections and discuss methodological and ethical concerns. Finally, I will go through the methodology and results of my research projects, before a broad discussion on the democratic value of comment sections.

The first topic covered in the synopsis is toxicity and anonymity. Due to concerns surrounding toxicity in comment sections, several news sites have closed their comment sections in favor of using their Facebook pages for engagement with their articles. Politico, one of the news sites studied in this thesis, is one of them. And among news sites that still use comment sections, more and more of them require commenters to identify themselves and post using their real names. One example of such a news site is the Norwegian newspaper VG. It is important to understand how these developments effect comment sections. If anonymity increases unwanted behavior, comment sections that require identification may serve a more positive role than those that do not. On the other hand, the ability to remain anonymous is thought to lower the threshold for participating in comment sections, and it is therefore important to understand if the requirement of posting with real names has a significant improvement on the quality of comments.

The second topic is how accusations of trolling in comment sections affect the debate. In recent years there has been an increasing awareness of trolling behaviors. The openness of comment sections could facilitate various sorts of trolling but could also make it easier for the users to identify and counteract such behavior. In this thesis I will investigate how commenters are affected by accusations of trolling by observing what happens when such accusations are made.

The third and final research topic is about criticism of the media in comment sections. It has been theorized that comment sections can be a valuable forum for constructive criticism of the media, and that commenters can serve as gatewatchers who improve the work done by journalists. Therefore, I have analyzed criticism of the media in comment sections to better understand how commenters can serve as gatewatchers.

1.1. The structure of the thesis

In chapter 1 I introduce the topic and research questions, and I look at various definitions of comments, comment sections and toxicity. In chapter 2 I present background information on technology and democracy, and on the readers and authors of comment sections. Chapter 3 presents literature about comment sections, toxic disinhibition, computer-mediated communication, and anonymity. In chapter 4 I will present democratic theories and how they can and cannot be used as a framework for discussing comment sections. This is important because a thesis about the role of comment sections in a **democratic** society should engage with **democratic** theories. The presentation in chapter 4 will include various democratic theories, some of which are closely related, to get a wide understanding of democratic thinking. We will see in this chapter that most democratic theories are not very useful as a framework when considering the democratic role of comment sections.

In chapter 5 I introduce my methodology. I also write about methodological and ethical challenges when researching comment sections and how I have chosen to solve some of these issues. In chapter 6 I discuss the results of my three research projects and the democratic theories presented in chapter 4. The role of comment sections in a democratic society will be explored through the results of the research projects and the democratic theories. In chapter 7, the conclusion is presented: that the role of comment sections in a democratic society is challenging, mostly because of the prevalence of toxic disinhibition. Despite this, comment sections have the potential for being a valuable platform for constructive criticism of the media, and they incentivize people to engage with the news media and serve as a public platform where people have their opinions challenged.

Finally, the three articles are presented in full at the end of the thesis.

1.2. Positioning the thesis

The current research project involves comment sections on mainstream news sites and human behavior in an online environment. Despite this, and despite it being necessary to reference sources from media studies and psychology, this thesis is positioned within the humanistic field of digital culture. For this reason, the background information will focus on the relationship between technology and democracy, not technology and the media or psychological research into human behavior in general.

Admittedly, comment sections are both a part of the traditional medium that is the newspaper, and a technological platform related to the internet and discussion forums. While media studies often have an interest in the media industry and would investigate the history of newspapers and the press, digital culture is more about the relationship between technology and culture ("Studer digital kultur" 2016). My topic of interest, as a researcher in the field of digital culture, are comment sections as a technology that exists within a culture.

Digital culture is also about the relationship between technology and humans, and I have an interest in how the technology of comment sections influence and is influenced by humans. In other words, I am interested in online human behavior – or spychology. Human behavior that is relevant online, might also be relevant in other environments. But being positioned within the field of digital culture, this thesis will only consult psychological references that are relevant to the technologies and online behavior covered by the thesis.

In summary, while the vast fields of psychology and media studies are relevant to this thesis, the background theme for this study is how technology influence society and human behavior. More specifically, how comment sections influence democracies as both a societal construct and the behavior of those humans participating in democratic institutions. Therefore, psychology and media studies sources will be referenced when relevant, but the focus of the thesis will remain within the field of digital culture.

1.3. Terminology

1.3.1. Defining comments and comments sections

For the purposes of this thesis, I define comment sections as a sub-category of forums found at the conclusion of publications such news articles, blog posts, images and videos, and are specifically designed to allow readers to respond to and discuss the content of the publications. Comments are user-generated posts on a comment section following an online publication. This definition of comments assumes that a comment is dependent upon being in a comment section for it to be a comment. Therefore, the definition of comment sections is more extensive, as it can be argued that qualities of comment sections, such as them being found at the conclusion of a publication, also will apply to the individual comments within them.

By its simplest and broadest definition, a comment is an observation or remark expressing an opinion or attitude, or a note explaining, illustrating, or criticizing the meaning of a writing (Merriam-Webster 2019a). Commenting goes back in history to ancient times, where complicated writing systems meant that readers required help deciphering texts. Therefore, the ancients developed conventions for annotating their works known as *scholia* (Reagle 2015, 23).

In the modern, digital world, commenting is usually referred to in the specific context of writing posts in a comment section, though this thesis is only concerned with comment sections on newspaper articles. Comment sections are defined by Artime as "forums attached to the conclusion of online news stories or blog posts and are designed to increase audience interactivity with the content contained in said stories" (2016, 1). This definition is not only helpful for defining what a comment section is. It also reveals something about the relationship between comment sections and forums, something that must be clarified for the purposes of defining comments. The forum, a word that can be traced back to the marketplace or public place of ancient Roman cities, can be defined as a public meeting place or medium for open discussion (Merriam-Webster 2019b). The online forum, then, is an online space for discussion. Forums have existed since the early days of the internet (Hubler and Bell 2003, 281; Gonçalves 2015, 1). They are online spaces specifically designed for discussion. Comment sections, according to Artime's definition, are forums. They are specifically designed for discussion among users, but they do not exist in a space of their own. Instead, they are attached to content such as news articles, blog posts, images or videos, and the topic of discussion is the content to which they are attached (with the exception of spamming and off-topic discussions).

This definition of comment sections as forums attached to content complicates the definition of comments themselves, as posts on a forum then can be regarded as comments. However, there is a distinct difference between forums and comment sections: the original content being commented on. In a forum, this content is a post by a user, and all replies to this original post are made by other users with the same rights and abilities to make their own original posts. All the users are equal and can be compared to a group of people in a room having a conversation with each other. Anyone can begin a conversation, and everyone is participating on the same level. In a comment section, however, the original content is quite different from the following comments. It can have a different modality and it is created in a different process than the comments – there is a technical and qualitative difference between creating a news article or a video on YouTube and commenting on it. The creator of the original content is not on the same level as the commenters, who can be compared to an audience listing to a speaker while making short comments on the content of the speech.

If comment sections are a subcategory of forums, then comments can be seen as a subcategory of forum posts. Comments can be regarded as a specific type of forum post that are posted on a comment section in response to an article, blog post, image or video. That does not mean that a comment must be a direct response to the original content, as commenters can respond to each other in the comment sections. But they are created in the public space at the conclusion of some content, where the content is being responded to and discussed by the commenters.

The American academic and writer of the book *Reading the Comments,* Joseph M. Reagle Jr. has an extensive definition of comments. He defines them as a genre of communication that is asynchronous, social, short, being written in context of something with a writer as a source and an audience, and being reactive, in that it follows as a response to and is found below a post, article or video (2015, 2 & 17). There are some problems with this definition, and not all his defining characteristics will be a part of my own definition. Firstly, I object to including asynchronicity. This characteristic does not help in separating comments from other forms of text-based computer-mediated communication, as asynchronicity is a common trait of all such communication – with the possible exception of chatting where all participants are present. There is also a problem with defining comments as short. While this may be the norm, there is no reason that comments cannot be longer. Furthermore, what is considered short is subjective and dependent on the context, making it difficult to use the term *short* in a definition. Finally, it is problematic to define a comment as being in response to something. This would imply a comment is related to some original post, which does not have to be true. There are no technical limitations on what a commenter can or cannot write. Moreover, as noted earlier, comments can be written in response to comments by other commenters.

Another researcher, Ian Rowe, describes comments as a feature that provides users with a public space at the end of each article in which they are invited to contribute their own opinions, perspectives and expertise to the content produced by professional journalists (2014, 122). This definition, however, I would argue is too specific. It specifies articles produced by professional journalists as a requirement for commenting, but comments can be written on all sorts of content online, from blog posts to YouTube videos.

1.3.2. Defining unwanted comments

Having defined what a comment and a comment section is, it is time to consider what term should be used for unwanted comments, a topic that will come up often in this thesis. Much research has been done on unwanted behavior online, including comment sections. There is, however, no single terminology to describe such behavior. Papacharissi (2004) created a coding scheme where online communication would be coded as uncivil or impolite. Uncivil is a term used by several researcher (Rowe 2014; Santana 2014), and is in part used in my own research. Because I use Papacharissi's coding scheme in my own research, it is unavoidable that I will at times write about uncivil and impolite comments, as these are the two main categories in the coding scheme. In general, however, I will mostly be using the term toxic comments when writing about unwanted comments. Toxic comments are those that display toxic disinhibition, a term coined by Suler (2005) to describe online behavior that is rude, critical, angry, hateful or threatening.[1]

[1] Suler also includes visiting places of perversion, crime and violence when defining toxic disinhibition, but this is not relevant in my own definition.

2. Background information

2.1. Technology and democracy

Democratic systems have taken advantage of technology for millennia. An early example of this is the Kleroterion, an intricate device in Athens in the 4th century BCE that was developed for making allotments – to randomly select who would serve on a jury or hold office (Rhodes 2012; Bishop 1970). As for public discourse, one early technology of interest (if one were to consider anything that has been invented or developed to be a technology) was the agora, the public center of ancient Greek cities serving as, among others, a political meeting place (Boehm 2012). Such a public meeting space allowed for the sharing of thoughts and ideas. But the agora had other important functions related to trade, religion and administration (Boehm 2012), as did most public spaces in the ancient world. But at some point, a public space for drinking was developed, called the tavern. At a time when all public spaces had some higher function - temples were for worshipping, markets were for trading, and so on - the tavern represented something new. As Steven Johnson writes: “The tavern was not a space of work, or worship; it was not a home. It existed somewhere else on the grid of social possibility, a place you went just for the fun of it” (2016, 220). Taverns, bars and pubs were not just revolutionary because they were a place of fun. They were revolutionary because they constituted a public space that was run by the people, for the people. In addition, they were a place of relatively free speech where people of different social standings could meet. The word *pub*, in fact, is derived from *public house* (Merriam-Webster 2019c), indicating its historical purpose as a place belonging to the public.

Bars, pubs and taverns have been seedbeds of social and political rebellion, such as the American Revolution and the movement for gay rights in the 1960’s (Johnson 2016, 222-225). These movements were not dependent on the existence of bars, pubs and taverns. However, as Johnson writes:

> American independence wasn’t *caused* by the prevalence of tavern culture in the colonies. There were many forces at work, some of them likely stronger than the space

> of dissent that the tavern offered the early revolutionaries. But the existence of that space was nonetheless a determinate factor in the way the events unfolded… so much of the debate and communication relied on the semipublic exchange of the tavern: a space where seditious thoughts could be shared, but also kept secret. (Johnson 2016, 224)

The idea of a semipublic space is closely related to Habermas' theory of the bourgeois public sphere, a topic that is very popular to discuss in media sciences (Lunt and Livingstone 2013). Habermas wrote about coffee houses and salons of the 18th and 19th century in much the same way that Johnson wrote about bars; they were centers of literary and political criticism where discussions were open for anyone to participate. Habermas' public sphere lay in the overlapping space between the private and the public; a semipublic space with open participation, a disregard for people's status and rational-critical debates independent from the authorities (Habermas 1991, 32-37). According to Habermas it was, among other things, the emergence of the press that would lay the foundation of the public sphere. Starting as a tool traders, capitalists and the authorities, the press developed to become more independent and focused on reasoning, knowledge and science (1991, 15-25). However, according to Habermas, the media has also caused the decline of the public sphere in the 20th century, as it has become refeudalized and commercialized (1991, 158-162). Beginning with theaters and concert halls in the 19th century, and later with radio and television, passivity and crowd silence became the norm. According to Sennett, this makes reasoning and debate between individuals almost impossible (1986, 282-283). This view, of course, assumes that individuals do not debate what they are passively experiencing with each other. One can argue, however, that people are still connected to the public sphere, by connecting their lives and homes to the public through the media technologies they engage with every day (Gripsrud 2017, 15). In this, what Gripsrud calls an imaginary common space, people gather information, have conversations and experience culture.

While communication technologies had improved for centuries - particularly after the invention of the printing press – most people could not easily broadcast their opinions to the masses. 20th century mass media had a special role in society as trusted one-to-many broadcasters. The development of 20th century media is closely tied to the United States, who James Curran claims is the principal originator and exporter of “a great media experiment” (Curran 2011). According to Curran, the American news media system is based on the idea that the media should be organized as a free market to be free from the government, and that it should serve democracy by being staffed by professionals seeking to be accurate, impartial and informative. This is contrasted with the cowed journalism of authoritarian states, the fusion of media and political power in Italy, and the tabloids of Britain. However, this ideal image of American journalism might not be accurate, according to Curran. He argues that American journalism is the product of an unequal society, and that the media helps to legitimize this inequality by sustaining the money-driven nature of American politics. In contrast to the American media, Curran reports that Danish and Finnish news media broadcast more news at peak times than in the U.S., and that Scandinavians are better informed because their news media covers more political and international news.

Because of the high cost of publishing, there was little room in the 20th for most people to share their thoughts and ideas to a wider public. Mass communication involved one-to-many broadcasting, with only a few communicators and a large, passive audience. Pubs, cafés and other public venues would continue to be the primary forums for most public discussions among the general public. But with the development of new electronic communication technologies, public discussions would find a new home and develop into the online comment sections and discussion forums we know today. Suddenly, information was no longer controlled by a few people with access to mass media, bringing hopes of a democratization of information. During the 1970’s, authors thought that the internet offered new possibilities of generating a public discursive and deliberative structure that could revitalize democracy and stimulate public debate and social change (Gonçalves 2015, 1).

Today's audience is not a passive one, as they can share and express themselves online. The audience engage with content on social media, newspapers facilitate online discussions and blogging (Rettberg 2014, 51), and comment sections have become practically an industry standard. It seems then, that the traditional media has attempted to integrate the previously passive audience, with questionable results as this democratization of publishing on news sites has brought with it what some would call a toxic or uncivil environment (A.A. Anderson et al. 2018; M. Knustad and Johansson 2021; Coe, Kenski, and Rains 2014). As I discussed in depth in my master thesis (M.A. Knustad 2018), there has been a trend in recent years of online publications closing their comment sections due to spamming and bad behavior by the commenters. This includes publications such as *The Chicago Sun-Times, Popular science, Reuters, The Week, The Verge* and *USA Today* (Bilton 2014; Ellis 2015; Finley 2015). In Norway, *Dagbladet,* closed their comment sections in 2016 (Ramnefjell 2016). To continue to facilitate public interaction and debate, publications that close their comment sections tend to focus more on their Facebook pages as a forum for debate. In addition to this, many publications who continue to offer comment sections on their articles use a comment section plugin by Facebook, raising concerns about privacy (Reagle 2015, 8-9). It would seem that the most recent development of comment sections, whether it is their closing or taking steps to require users to identify themselves, is driven by concerns surrounding the civility of comments.

The latest developments of removing anonymity or closing comment sections to combat toxicity almost suggests that there is a deterministic view of the technology of comment sections. Technological determinism is “the idea that technology develops as the sole result of an internal dynamic, and then, unmediated by any other influence, molds society to fit its patterns” (Winner 1980, 122). Within this framework, comment sections developed, and then, because of the toxic nature of comment sections, they must now be closed. Even the removal of anonymity in those comment sections not being closed suggests a deterministic view in which there is no solution to toxicity but to change the technology. In contrast, social determination of technology does not focus on the technology itself, but the social or economic system in which it’s imbedded (Winner 1980, 122). Through this framework, comment sections are not of interest, but instead one should focus on how and why people use them, including for writing toxic comments. There is no point in making changes to the technology itself because it will still be used by the same people in the same social setting.

Both technological determinism and social determination of technology could be criticized for being too simplistic, so Winner provides us with a third alternative for how to view technologies where it’s not just one or the other (Winner 1980). Technological objects, or artifacts, have political properties and they embody some form of authority. As an example, Winner points to the extraordinary low bridges over the parkways of Long Island, New York. The bridges were designed to be too low for buses, thereby ensuring that only those rich enough to own a car could pass. But artifacts may also have unintentional properties, such as when public spaces in the U.S. were not made accessible to the handicapped before the 1970s (Winner 1980, 125). Technology, then, is influenced by people and society, but it also is an influence on people and society. And if Winner is correct, the same should be true of comment sections. Therefore, we should expect to see ways in which comment sections affect people and society, but also how people and society affect comment sections.

2.2. Comment sections and the media

With the implementation of the world wide web in the 1990's, some newspapers began to publish text-based stories online. But after the release of the first graphical web browser in 1994, full online editions were created. By the year 2001, the number of online newspapers in the U.S. alone had reached over 3.400 (Li 2010, 1-2). In Norway, Brønnøysund Avis became the first online newspaper on March 6. 1996, closely followed by the national newspaper Dagbladet two days later (Solheim and Syvertsen 2021). Alongside the development of online newspapers, comment sections have grown in popularity to become almost an industry standard. A content analysis by Stroud, Muddiman, and Scacco (2016) found that 90% of online news sites had comment sections. Another study of the 150 largest American newspapers found that 92% of them had comment sections (Wallsten and Tarsi 2015, 1023).

In a survey reported on by the Center for Media Engagement (Stroud, Van Duyn, and Peacock 2016) 55% of Americans reported that they had left an online comment, of which 77.9% had done so via social media. 77.9% of Americans reported having read a comment at some point. When looking at comments on news articles specifically, 50.7% of Americans reported never having read or written a comment on a news article. Of those who did comment on news articles, 53.2% did so monthly or less frequently. It seems then that only a quarter of Americans frequently comment on news articles. News commenters are more often male, have lower education and a lower income than those who only read comments. In Norway, a 2019 report found that 28% of participants reported having participated in online debates, of which only 8% had debated in comment sections (Medietilsynet 2019).

One study found that as much 84% of newsreaders read comments attached to news articles, and that comments can have a significant effect on readers' perception of public opinion, and even change their personal opinions (Toepfl and Piwoni 2015, 467). People who participate in discussions in comment sections are often seen as someone with serious interpersonal and intellectual problems (Artime 2016, 2). However, Artime's analysis of demographic data from the Pew Research Center gives us a more accurate view of who comments on comment sections (2016). Artime studied data from 2008, 2010 and 2012 about how many Americans have contributed to comment sections. In 2008, 11% of respondents reported having commented on online comment sections. Men were more likely to comment than women were, and unmarried and unemployed people were more likely to comment than married and employed ones. This means that unemployed, unmarried men were the most likely to comment. This trend continued in 2010, even though the number of respondents reporting to have commented on online comment sections had increased to 24%. From 2010 to 2012 there were significant changes in the demographics of people commenting. In 2012 gender, marital status or employment were no longer good predictors of whether someone commented on online comment sections. Age, educational level and race, however, were. Young people, highly educated people and white Americans were more likely to comment. There was also a correlation between commenting on comment sections and offline political activity (this data was only available for 2012, making it impossible to say if this was also the case in 2008 and 2010).

A 2020 report (Stroud, Murray, and Kim 2020) states that when comments were turned off on news articles, about 75% of those who had previously commented on a news site didn't notice that the comments had been removed. This suggests that commenters aren't always actively seeking out the comment sections. The 25% who did notice, however, said that it "made the experience worse". Another finding in the report was that the average time spent on the site was lower when comments had been turned off. This would make sense, considering commenters would spend less time there when they would not be reading or writing comments. Furthermore, this may hint to one reason why news sites host comment sections, as more time spent on their sites could translate into more views and clicks on advertisement. As Williams and Sebastian (2022) point out, by closing comment sections, news sites risk losing readers to social media sites. In the end, the fear of losing readers might be the biggest reason why comment sections still exist on professional news sites.

3 Previous research

3.1. Comment sections as forums for participatory journalism

Comment sections have been around since the mid 1990's, and are now seen as a staple of online news sites (Artime 2016). At first, journalists responded to comment sections with caution, and were skeptical about the quality and trustworthiness of user-generated content (T. Graham and Wright 2015; Toepfl and Piwoni 2015). Some researchers, however, have argued that comment sections are a form of participatory or constructive journalism (Løvlie 2018), and newspaper editors view comments as one of the most successful forms of audience interaction (Singer, Paulussen, and Hermida 2011). A study by T. Graham and Wright (2015) found that journalists were held accountable by the readers, which many of them felt improved the quality of their work. Journalists reported that they read roughly the first fifty comments on their articles. Some argued that comments made them reflect on what they wrote and how they wrote, keep paper trails of their stories, and that they received new stories and leads from comment sections. There was, however, little evidence of debate between journalists and readers. This was largely explained by a lack of time, but also fear of personal attack (T. Graham and Wright 2015).

The possibility of comment sections being a form of participatory or constructive journalism is intriguing. News media serve an important function in a democratic society as they can investigate public figures and organizations and hold them accountable. Journalists and editors can make mistakes, though, and commenters could serve as gatewatchers (Bruns 2005, 17-18), a term used by Rettberg (2014, 108) when writing about how bloggers can check the output of newspapers and point out mistakes. Commenters can also make constructive critique about the contents of news articles and can provide new leads to the journalists. In other words, there is the potential for commenters serving as valuable critics of the media, as long as their critique is constructive. Therefore, it is important to understand how the media is criticized in comment sections, which is the focus of one of my articles and the discussion in chapter 6.3.

3.2. Toxic disinhibition in comment sections

Since the 1990's academics and psychologists have attempted to explain antisocial behaviors online, often blaming the anonymity that the internet provides for the tendency of seemingly normal people to show disinhibited and toxic behavior online (Suler 2005; Lapidot-Lefler and Barak 2012; Gonçalves 2015; Stroud, Muddiman, and Scacco 2016; Rowe 2014). Phillips (1996) explored how a newsgroup used flaming as a defensive measure when faced with difficulties from new members who were challenging established norms. John Suler developed theories about why people behave badly online (Suler and Philips 1998), and explored theoretical explanations for what he calls the online disinhibition effect (2005).

One of the criticisms against comment sections is that comments are often uncivil and impolite, with a lot of disagreements and arguments. This can scare people from voicing their opinion in comment sections, a problem that is more prevalent with minorities (Rossini 2019). 33.9% of news commenters and 40.9% of comment readers reported argumentative comments as their reason for avoiding writing or reading comments (Stroud, Van Duyn, and Peacock 2016). In other words, a substantial number of people will not participate in comments because of perceived arguments. Incivility then, it seems, does have an impact on comment sections, a topic I will explore further in my article about anonymity in comment sections.

There have been many attempts at defining uncivil comments. Papacharissi developed a coding scheme where uncivility in online discussion forums was defined as being a threat to democracy, denial of people's personal freedoms, and stereotyping social groups (2004). Ian Rowe used this coding scheme in his research on the effects of anonymity on civility in comment sections (2014). He found that 6% of comments on the Washington Post comment section were uncivil. Other researchers have found varying numbers of uncivility in comment sections. In a literature review, Vergeer found that the number of uncivil comments reported by researchers vary from 4-22% (2015, 746). These differences may be explained by different definitions of uncivility, as well as differences across studied platforms and websites. Uncivility in comment sections has also been described as dark participation, which refers to different forms of user engagement by wicked actors driven by sinister, strategic, tactical or "pure evil" motives, attacking despised targets directly or indirectly with the aim of manipulating different audiences (Frischlich, Boberg, and Quandt 2019; Quandt 2018).

Nevertheless, the number of toxic comments reported by researchers indicate that the great majority of comments are civil. Graham and Wright (2015) found that the comments they studied were deliberative, and that discussions were typically rational, critical, coherent, reciprocal, and civil. They also note that it appeared that the participants held a wide range of political views. Another uplifting finding was that user-generated information was routinely challenged and debated by other participants. In doing so, some of the fear journalist have about the spreading of misinformation, may be relieved.

One form of incivility in online forums is trolling, which will be further explored in one of my following articles and in chapter 6.2. Trolling is defined by Buckels, Trapnell, and Paulhus (2014) as "the practice of behaving in a deceptive, destructive, or disruptive manner in a social setting on the Internet with no apparent instrumental purpose". Originally, the term “troll” referred to jokesters who behave in an antagonistic way online for their own amusement's sake (Hardaker 2015, 202), and are motivated by boredom, attention seeking, revenge, pleasure and a desire to cause damage to a community (Shachaf and Hara 2010). Trolling can also refer to when someone is engaging in large-scale harassment campaigns, impersonating multiple identities, and engaging in extremist activities (de Seta 2018, 392). In recent years, trolling has also been widely used when referring to political influence through social media. Another concept that has received attention in recent years is “bots”, which are computer programs used to add content in social media platforms. While these bots can have a wide variety of benign usages, such as user interaction and the automation of tedious tasks (Lebeuf, Storey, and Zagalsky 2018), they can also be used to spread disinformation (Broniatowski et al. 2018; Bastos and Mercea 2019).

3.3. How computer-mediated communication affects us

There are many things that can be thought to influence how we communicate online, and many possible causes for both prosocial and antisocial online behavior. In this chapter I will cover many of these, focusing mainly on what might cause people to behave in an antisocial way online.

Computer-mediated communication has been a true revolution in human history, as described by psychologist John Suler (2016, 2):

> Just yesterday, comparatively speaking in the many millennia of our evolution, we humans did something quite remarkable. We created an entirely new environment for ourselves, one that intersects but also transcends the physical world as we have known it for all these hundreds of thousands of years. People call this new digital realm "cyberspace."

When using the Internet, people tend to use spatial metaphors like "worlds", "domains", and "rooms" to describe online environments. And when moving around on the web people describe the experience as "going" someplace (Suler 2016, 22). This indicates that we see the internet as a real place when speaking of it. There are, however, important differences between the online world and the real world that affect how we communicate in face-to-face communication and computer-mediated communication. Sociologist and social psychologist Erving Goffman compared face-to-face interactions with a theatrical performance, where a person interacting with someone is described as playing a role and offering a performance (Goffman 1956, 10). This comparison is interesting when considering the implications of computer-mediated communication. If one were to extend the metaphor of communication being a theatrical performance, is the online world another stage to perform on? Or is the online world, due to its non-corporeal nature, a less real stage within a stage? If face-to-face communication can be seen as a performance, it is certainly reasonable to argue that computer-mediated communication is a performance as well. But with less information about the performer, the audience may have greater difficulties interpreting the performance.

Any form of communication requires that the individuals communicating have information about each other. The audience tries to access information about the communicator by getting information about him or her. This information, according to Goffman, helps to define the situation, enabling others to know what the communicator will expect from them and what to expect of the communicator. Many sources of information are available, and if the communicator is unknown, the audience can get clues from his or her conduct and appearance (Goffman 1956, 1). In the online space, however, clues about the communicator are limited. Moreover, the individual communicating also has little information about the audience. According to Marwick and Boyd, every participant in a communicative act has an imagined audience. We understand that the audience on social media platforms such as Twitter or Facebook is potentially limitless. But because of our limited understanding of the social media audience we often act as if there is a clear limit to the audience (2010). This may result in our online communication not being tailored to the actual audience.

The differences between message, meaning and context has been theorized to be an important part of communication (Smith 1965). A message is encoded by a communicator and decoded by the recipient (Hall 2006), which may lead to misunderstandings. In face-to-face communication, we take advantage of cues to decode a message. Context, tone of voice and body language make up a significant portion of available ques. Albert Mehrabian found that only 7% of a message is communicated verbally, while 38% is communicated through vocal elements such as tone and pitch, and 55% is communicated through body language and facial expressions (Mehrabian 1971). While Mehrabian's theories have received some criticism (Casselberry 1971), the importance of non-verbal communication is recognized by most researchers. Especially when communicating emotions, non-verbal cues have been found to be important (Gilovich et al. 2016).

In computer-mediated communication, non-verbal cues that we normally use to help decode a message are lost, and only the words written by the communicator are left. While one would think that this would lead to fewer misunderstandings, it does allow for more freedom in interpreting what other people mean. Irony has been found to be especially difficult to successfully transmit in an online environment (B. Graham 2016; Kruger et al. 2005).

3.4. Anonymity

Much research and media attention has been focused on the negative impact of user-generated comments. Anonymity, "the condition in which a message source is absent or largely unknown to a message recipient" (Scott 2012, 128), has often been used to explain toxicity in comment sections. In most cases of anonymity, whether they be Victorian authors hiding their gender or online commenters sharing their political views, a pseudonym is used. There are two types of pseudonyms: 1) Those where the source of the message is perceived as fictitious, such as when using a clearly made-up pseudonym such as *California42* or *MrFantastic,* or any other of the many imaginative pseudonyms that can be found in online forums, comment sections and chat rooms. 2) Those pseudonyms where the source of the message is perceived as factual (Scott 2012, 129). The latter type of pseudonym is especially problematic when doing research into anonymity using examples from online sources, as any name that seems to be factual could potentially be a pseudonym.

There is a long tradition of people attempting to be anonymous, as there have always been reasons for people to want to hide their identities. *The Federalist Papers,* published over two hundred years ago in the newly independent United States, were published under the pseudonym *Publius*, and anonymous sources such as *Deep Throat* during the Watergate scandal has been an important resource for newspapers for decades (Scott 2012, 127). During times of anti-Semitism, Jewish authors and artists adopted less Jewish names, and during the Victorian era, female authors used male pseudonyms out of fear of being dismissed based on their gender (Hogan 2013). Even in modern times, female authors have found it necessary to hide their gender at least partially. The most prominent example of this is Joanne Rowling, who published the Harry Potter books under the name J.K. Rowling. This was a marketing ploy because boys tend to not read books written by female authors (Savill 2000). While the example of J.K. Rowling may not be a clear example of a modern author using a pseudonym not connected with her real name, it is nevertheless an example of someone seeing the need to hide a part of her identity.

As I will discuss further in chapter 6.1., the above arguments in favor of anonymity could also be arguments in favor of allowing anonymous comments. At the very least, we must question how effective removing anonymity is at reducing toxicity, if we by removing anonymity raise the bar of participation. In the past few years, there has been a movement against anonymous online comments. *The no anonymity movement* started in 2010 and is motivated by two assumptions: anonymity leads to hostility and insults, and anonymous comments are thought to exert a strong influence over internet users by allowing for cyberbullying (Wallsten and Tarsi 2015, 1019-1020). 47.9% of writers and readers of comments in a survey thought that allowing anonymous comments raises the level of disrespect (Stroud, Van Duyn, and Peacock 2016), suggesting that a large portion of the users of comment sections see anonymity as a problem to some extent. However, when the Norwegian newspaper *Aftenposten* removed the possibility to be anonymous in their comment sections in 2013, most commenters were critical of the move, arguing that people with controversial opinions will not want to come forward, resulting in a loss of diversity (Elgesem and Nordeide 2016).

To counter uncivil comments many news sites in recent years have adopted comment sections that use a Facebook plugin, where users must log in to their Facebook accounts to comment. Some sites have closed their comment sections and are instead using their Facebook pages as a platform for users to debate news articles (Bilton 2014; Ellis 2015; Ramnefjell 2016). This raises concerns about privacy (Reagle 2015, 8-9) and moderation. All platforms must to some degree moderate user generated content by certain criteria. On commercial platforms, such as news websites or social media sites, these criteria are set by a company who must find a way to make profit, and reassure advertisers and investors (Gillespie 2018). By handing their comment sections over to Facebook, news sites are allowing Facebook, to some degree, to both track their readers and moderate content. The existence of anonymous comments may even affect how readers of an online newspaper rates the media itself. Wallsten & Tarsi found that the average rating internet users gave the media suffers when anonymous comments are included alongside news reports (2015, 1031).

One of the leading opponents of anonymity is Meta (Facebook) founder and CEO Mark Zuckerberg, who has built his company and services around the concept of people using their real names online. Zuckerberg has been quoted saying, "having two identities for yourself is an example of a lack of integrity" (Kirkpatrick 2011, 199). This claim was countered by Christopher Poole, the founder of the infamously anonymous website 4chan, who said, "Zuckerberg's totally wrong on anonymity being total cowardice. Anonymity is authenticity. It allows you to share in a completely unvarnished, raw way." (Halliday 2011).

This disagreement between Zuckerberg and Poole reflects the fact that different platforms situate themselves differently along what Suler calls the reality dimension. The reality dimension is evaluated by asking if the experience of an online domain is based on imagination and how much it is grounded in the familiar everyday world (Suler 2016, 46). Online games and creative spaces often encourage fantasy and creativity, while social media platforms encourage, and often require, people to be who they really are. It is problematic to state that anonymity in general should or should not be allowed, as it is dependent on the platform, its uses and its social norms. Zuckerberg, as the CEO of Facebook/Meta, is well within his right to situate his platform on the real, non-anonymous side of the reality dimension. But other platforms, such as 4chan, that are used in a very different way, may do the opposite.

Zuckerberg's statement about having two identities for yourself being an example of a lack of integrity is also problematic because it disregards the fact that anonymity is considered a right under the first amendment in the United States (Scott 2012). Zuckerberg also seems ignorant to the fact that having multiple identities is a part of the social human experience. The most basic example of this is the private and casual identity versus the professional identity. One can also differentiate between the identities we show our close friends, our families, casual acquaintance or people with whom we share a hobby or interest. Kirkpatrick notes that up until now it has been possible to share one or the other identity depending on the social context (2011, 199), something that becomes impossible on Facebook where we are required to use only one identity. Marwick and Boyd (2010) argue that pseudonyms make it possible to avoid what they call context collapse, where multiple audiences are flattened into one, which makes it difficult to differ self-representation strategies. In addition, having multiple identities, or personalities, is considered a natural part of social interaction. Sociologist and psychologist George H. Mead noted in the 1930's that a person's attitudes and gestures were influenced by the social situation (1934). According to role theory in social psychology, we juggle our various social roles, making multiple personalities something that is a normal part of human nature. In this sense, our online lives is just another example of this social juggling act which requires the use of multiple personalities or identities (Suler 2016, 73).

It has also been argued that anonymity facilitates free expression, the sharing of unpopular ideas, whistleblowing, obtaining sensitive information such as in research, focus on the message rather than the messenger, protection from subsequent contact, avoiding persecution and encouraging innovation and experimentation (Scott 2012). 66.6% of comment writers and readers agreed with this sentiment in a survey, reporting that allowing anonymity in comment sections allows participants to express ideas they might be afraid to express otherwise (Stroud, Van Duyn, and Peacock 2016). And the idea that anonymity automatically leads to uncivility may be too simplified. When studying group processes, it was found that group members who were anonymous made more comments, were more critical and probing, and were more likely to embellish on group members' contributions (Jessup, Connolly, and Galegher 1990). Visually anonymous people have also been found to be more willing to disclose information about themselves (Joinson 2001; Chiou 2006), though such findings have been put into question by a later study with opposite findings (Misoch 2015).

Bernie Hogan argues there are several reasons why someone would choose to be anonymous (2013):

- **External pressures** may motivate an individual to mask his or her real identity in order to be treated in a desired way. As an example of this, Hogan points to female Victorian writers using male pseudonyms out of fear of being dismissed based on their gender, as well as Jewish authors and artists that adopted public names that appeared less Jewish.
- Another motivation for being anonymous is **internal motivations,** where an individual has a desire to adopt a different persona. Again, Hogan points to historical writers such as Mark Twain who used a pseudonym on some works to distinguish them from titles that were more serious.

- **Functional motivation** is the use of pseudonyms because of practical reasons, such as when naming a band, pope or royal person, or simply coping with the fact that other individuals may share your name. One example of this that most people online have experienced is when choosing an available e-mail address or having a short and catchy Twitter handle. Some individuals may also have to choose a western name online on services that require the use of Latin characters.
- **Situational motivations** may arise when posting things with one's real-life name makes it possible for two completely different posts from different sources are presented in the same search results. Sometimes a person will want to use a pseudonym for some topics and their real name for others to avoid such linking.
- **Personal motivations,** such as the desire to creatively escape their everyday life, can be a final motivation for the use of anonymity.

Hogan argues that we should not consider a scale moving from real names through pseudonyms to anonymity (2013, 293), and suggests that anonymity is a binary state – either you are anonymous or you are not. He makes the distinction between anonymity, a state implying the absence of personally identifying qualities, and pseudonyms, which he considers a practice. He may be right in this distinction, but when studying how people behave online one should also consider the consequences of behavior that may invoke a negative reaction from others. As Christopher Poole said, "The cost of failure is really high when you're contributing as yourself" (Halliday 2011). One may think that this cost is not too high when someone is using a pseudonym. However, some individuals in certain online communities, such as Reddit or in an online game, may have spent a lot of time and effort building the ethos of their online persona. Therefore, the consequences of social failure, mainly the negative reputation one's pseudonym might attract, may lead people who have invested a lot into their online persona to comply with social norms, despite not being personally identifiable.

3.4.1. Anonymity and toxic behavior online

The idea behind anonymity leading to toxic behavior is a simple one: When people think their identity is hidden, they feel less vulnerable about displaying behavior that would otherwise be suppressed. Not being accountable for one's actions seems to underlie most concerns about anonymity (Stein 2003; Scott 2012; Jacobsen, Fosgaard, and Pascual-Ezama 2018). Hirsh, Galinsky and Zhong suggests that disinhibited effects of anonymity, social power and alcohol intoxication emerge from a common psychological mechanism; lower activation of the Behavioral Inhibition System, resulting in that the most salient response is expressed in any situation, without regarding any prosocial or antisocial consequences (Hirsh, Galinsky, and Zhong 2011). Furthermore, the fact that others are anonymous may also lead a person to becoming toxic (Suler 2016, 99). Postmes et al. found that anonymous group members are more likely to be affected by social influence (Postmes et al. 2001). This would suggest that anonymous internet users are more likely to behave in an uncivil manner when others are uncivil.

Several studies have found that civility decreases when comments are anonymous. Ian Rowe found that there were more uncivility in comments on the Washington Post comment section than on the same articles on Facebook (2014). Rowe explains this difference with the fact that users of the Washington Post comment section are anonymous. This explanation, however, disregards other possible differences between the two platforms. Another study researched comments from different newspapers, some of them allowing for anonymity while other required commenters to use their real names, and found a significant relationship between anonymity and uncivility (Santana 2014). Lapidot-Lefler and Barak found that anonymity influenced the numbers of threats made by research participants, indicating that while anonymous, people may be more prone to threaten others. Anonymity was, however, not found to influence self-reported flaming, negative atmosphere or flaming-related expressions (Lapidot-Lefler and Barak 2012). Another study looked at the relationship between anonymity and cyberbullying and found that the more people feel that they are anonymous, the more likely they are to cyberbully others (Barlett, Gentile, and Chew 2016). Zimmerman and Ybarra found that anonymous participants in their study were more aggressive than those who were not anonymous (2016), adding to the evidence that anonymity may be a cause of uncivil behavior.

Other studies have not found a decrease in civility due to anonymity. Bae (2016) found that anonymity led to a greater feeling of in-group similarity and more attitude change, but less flaming and fewer critical comments. Janne Berg (2016) studied the effect of issue controversy and found it to have a greater impact on discussion quality than anonymity. However, there is certainly more evidence in the literature that suggests a correlation between anonymity and uncivil behavior. In addition, my own study presented in the paper *Anonymity and Inhibition in Newspaper Comments* (M. Knustad and Johansson 2021), found that while the relationship between anonymity and toxicity is real, it is very small and cannot explain all toxic comments.

3.5. The online disinhibition effect

If anonymity alone cannot explain why some people act in an uncivil or impolite manner online, other explanations must be considered. There are many possible factors that can influence our online communication. John Suler suggests several explanations for what he calls the online disinhibition effect, which is defined as benign or toxic uninhibited online behavior (2005, 2016). In addition to anonymity, which was covered in the previous subchapter, Suler suggests that asynchronicity, solipsistic introjection, dissociative imagination, attenuated status and authority, and invisibility may explain uninhibited behavior online.

- **Asynchronicity** in computer-mediated communication removes the constant feedback-loop of face-to-face communication, whereby people can self-regulate based on the feedback they receive from the person their communicating with.
- **Solipsistic introjection** refers to when reading someone's message might be experienced as a voice within one's head. The online person one is communicating with becomes a character within one's intrapsychic world – a character shaped partly by how the person presents their self.

- **Dissociative imagination** refers to the experience of the character one creates of oneself and others existing in a different world. Online actions may become a game, detached from the real world.
- **Attenuated status and authority** refer to how status and authority may disappear in an online setting because of the lack of real-world authority-cues.
- **Perceived privacy** is a term applied to how secure people can feel when they reveal personal information about themselves during professional or official transactions online. People experience themselves as being in a private encounter with online companions, even if they should know better.
- **Social Facilitation** refers to how the social environment can reinforce, amplify or fail to counteract the disinhibition effect. In some online environments, the actions of an uncivil or impolite person might be entertaining to the audience, who either encourage it or passively observe the hostilities. Even if someone wishes to interfere with toxic disinhibited behavior, they may not do so because of the by-stander-effect, also known as diffusion of responsibility. This phenomenon, where the presence of other bystanders at emergencies reduces the likelihood of someone helping, occurs because each bystander tends to assume that others will intervene, and thus each person feels less responsibility for providing assistance (Gilovich et al. 2016, 531).
- **Invisibility** refers to the fact that people communication in an online behavior cannot see each other, whether they are anonymous or not. Invisibility seems to be preferred by people when communicating online, also when communicating with people we know. Even though different types of video calls have been available for decades, and most smart phones today have this functionality, they are rarely used outside of a professional setting (Suler 2016, 18). Invisibility as a possible explanation for uncivility is in part supported by Lapidot-Lefler and Barak's study, where invisibility was shown to produce a more negative atmosphere (2012). Lack of eye contact was shown to have the highest effect on participants of the study. However, I would argue that lack of eye contact is related to invisibility, as eye contact is an important part of observing others

and the feeling of being observed. Anonymity was found to influence the number of threats being made by participants.

3.6. Social influences on how we behave online

People affect each other's behavior when communicating. Conformity, which is the changing of behavior or beliefs in response to real or imagined explicit or implicit pressure from others (Gilovich et al. 2016, 305), is a powerful influence on human behavior. In uncertain situations, conformity is more likely to influence our behavior, because we look to others, thinking that they must be better informed about the situation. Direct influence by others is the result of someone directly trying to influence others through persuasion. Indirect influence occurs when a person is affected by the available information about the behavior of other people (Cheng et al. 2015).

As I discussed in my master's thesis (M.A. Knustad 2018), the first people commenting on an article are not likely to be affected by social influence, because there is no previous behavior to conform to – at least not in the comment section of the article in question. Later commenters, however, have a lot more information about how others communicate and are more likely to adopt an established theme of commenting. Cheng et al. found that people on online bulletin boards conform by adopting both positive and negative information (2015). Social influence is also linked to aggression in commenting. If peer comments are aggressive, a commenter is more likely to write aggressive comments (Rösner and Krämer 2016). So, if an article has a lot of antisocial or uncivil comments, a newcomer is more likely to conform to this style of commenting.

If conformity and social influence can influence how people communicate online, it is not unreasonable to ask if this cannot also make people communicate in a more civil manner. While social influence does not solely effect people in a negative way, there are other factors that make social influence more likely to lead to uncivility in comment sections. Frequent contributors to comment sections have been found to be less civil and informational (Blom et al. 2014), meaning that those who comment the most are more likely to be uncivil. And because those who comment the most are the same people who are most likely to begin a discussion in a comment section, they are the ones who will set the tone for the following debate. If the first commenters on an article are more likely to be uncivil, social influence may cause additional commenters to behave in a similar manner. However, this is just speculation, and Blom et al.'s findings are challenged by the findings of Coe, Kenski, and Rains (2014), who concluded that frequent contributors to comment sections are more civil than infrequent contributors.

4 Comment sections and democracy

The topic of this thesis is the role of comment sections in a democratic society. To investigate this topic, one must consider what is meant by a democratic society by looking at various theories of democracy. However, democracy is not easy to define. Crick describes democracy as three different stories (2002, 236);

> *There is democracy as a principle or doctrine of government; there is democracy as a set of institutional arrangements or constitutional devices; and there is democracy as a type of behavior.*

According to Crick (2002, 323) we should not conclude that there is a 'true democracy'. Some may define democracy as 'majority rule' or the process of electing political leaders, some may equate democracy to *liberty* or *equality,* and some may define it as a complex political system. Plato attacked the idea of democracy in ancient Athens because he saw it as the rule of the ignorant over the knowledgeable, while Aristotle defended democracy, arguing that good government was the ruling of the few with the consent of the many (Crick 2002, 342-344). In more modern times, the term has evolved from the liberal ideas of the French revolution, focusing on how everyone has a right to speak their minds in matters of public concern, to the idea that all can participate if they care, "but they must then mutually respect the equal rights of fellow citizens within a regulatory legal order that defines, protects, and limits those rights" (Crick 2002, 362). Modern democracy, then, is usually seen and practiced as the idea of majority rule and the guaranteeing of individual rights. But that does not mean that there is a consensus on what democracy is, what it should be, and if it is currently being practiced.

At first glance, internet technologies that allow for communication and public expression are democratically valuable. Freedom of speech is a fundamental right in liberal democracies, and so it follows that any technology that facilitates the ability to express oneself plays an important role in a democratic society. This assumption, however, must be investigated further by looking not only at democratic theory, but how communication technologies such as comment sections compare to those theories and their ideals.

The early pioneers of the internet hoped that a new vitalization of democracy would take place as people connected digitally (Gonçalves 2015). It is difficult to say if the internet has been a democratizing force, or if public debate has improved because of it. This question is also too broad for this thesis, and I will focus on the democratic properties of commenting on news articles.

What is certain, is that people have never had such opportunities to share their opinions and participate in public debates as they do now. This includes comment sections. Therefore, it is tempting to say that comment sections have an important role in modern democracies – or at the very least, that it has the potential to be so. However, as we have seen, comment sections also allow people to act in an antisocial manner, which may threaten their potential as a democratically valuable tool. It is difficult to define what makes a comment or online discussion democratically valuable or find a way to measure the quality of commenting on a platform. In looking back to the early days of the internet described above, a goal of commenting might be found: to revitalize democracy and stimulate public debate. So, what are the qualities of democracy and public debate in an online world? Janne Berg argues that high-quality online discussions are characterized by rational reasoning, posting on-topic, and reciprocity and respects, and that participants of such discussions give arguments for their opinions, stick to the topic, show signs of respect towards others, and that they engage in dialogues rather than monologues (Berg 2016, 38). While Berg's view may be a good description of the preferred qualities of comments, comment sections could also be put in a framework of democratic theories to better be understood as a democratic tool.

Within the aggregative view of democracy, political preferences are taken as a given, and requires no justification. Philosopher Jeremy Bentham believed that the aggregation of individual votes, when cast anonymously by people voting according to their own interests, will reflect the public interest (Shafe 2014). The goal of aggregative democracy is to combine preferences in ways that are efficient and fair. Governments should make decisions either by putting political questions to a vote or by having political officials take note of the expressed preferences of the people, but put them through an analytic filter (Gutmann and Thompson 2004). The main focus of aggregative democracy is the act of voting; elections enable a society to make social choices when there is conflict between individual preferences (Perote-Peña and Piggins 2015).

Aggregative democracy could be argued to be related to representative liberal theory, a group of democratic theories that focus on representation through political parties (Ferree et al. 2002). In the realist school of democracy there is a belief that ordinary citizens are poorly informed, have no interest in public affairs, and are not well equipped for political participation. Representative liberal theorists believe that it is the main role of the citizen to choose which among competing political parties or politicians should have public authority (Ferree et al. 2002). Related to aggregative democracy is competitive democracy, which also focuses on elections. According to this theory, political elites act, whereas the citizens react (Strömbäck 2005). Procedural democracy, which is described by Strömbäck (2005) as a normative ideal and the minimum requirements a country has to fulfill in order to be democratic. This model of democracy states that citizens and politicians are expected to respect the rules and procedures of democracy and focuses on basic requirements such as the right to vote and freedom of expression.

Comment sections cannot be said to have much of a role in these theories of democracy. They do not have a clear role in the aggregative view of democracy or competitive democracy, as these philosophies concerns themselves mostly with elections. Elections are a method of measuring the preferences of the people, and comment sections could be another such method. However, this is a very ineffective method to do such a measurement. Firstly, it would require time and resources to extract the preferences of the people from the countless comments across different comment sections. Secondly, it is unreasonable to expect that all political preferences are accurately represented in comments that are written by only a small percentage of the people. Comment sections can potentially be of value for the media, as it provides a direct line of feedback to journalists. Journalists have been found to be affected by comments to perform better (Artime 2016), but other studies have shown that news organizations have not fully embraced such audience feedback. Comment sections are seen by news websites as not much more than an opportunity for their readers to debate current events (Domingo et al. 2008).

Representative liberal theory is, according to Ferree et al. (2002), about closure. While people have a right to their opinions, implying respectful disagreements, once a decision is reached, there is no need for further debate. Comment sections, however, do seem to be lacking in respectful disagreement, and there is no point of closure. A discussion in a comment section only ends when all commenters have said what they wanted to say, at which point the debate dies down on its own without anything that can be called *closure.* Within procedural democracy, comment sections could be argued to play an important role. This democratic theory emphasized freedom of expression, and comment sections represent a platform for free expressions of opinions. However, comment sections are not the only platform available for people to express themselves and cannot be said to be critical. Furthermore, toxic comments in comment sections challenge their role as a valuable forum for expression, as such toxicity could make people hesitant to participate.

Comment sections may play more of a role in democratic theories such as deliberative democracy or discursive theory because these theories emphasize open participation, and places high value on reasonable argumentation. According to Gutmann and Thompson, deliberative democracy has a goal of reaching a conclusion based on mutually acceptable reasoning among free and equal citizens (2004). Deliberative democracy is based on the goal of rational discussion, which can cause people to reflect on their opinions and judgements, producing unanimous preferences, thereby making the problem of social choice trivial (Perote-Peña and Piggins 2015, 94). Unlike the aggregative view of democracy, where the focus is put on the voting, deliberative democracy concerns itself with a dynamic process of open debate, also outside of an election cycle.

Central to deliberative democracy is the Habermasian public sphere, presented in the book *The Structural Transformation of the Public Sphere* by German sociologist and philosopher Jürgen Habermas (1991). According to Habermas, the public sphere grew out of the new administrational class of jurists, scholars, pastors and doctors called the bourgeois of the 17th century. Coffee houses and salons, where peoples' status and class were disregarded, became centers of literary and political criticism where discussions were general and open for anyone to participate. The public sphere is influenced by the state of the media. It was, among other things, the emergence of the press that lay the foundation of the public sphere. The press began as a tool for traders and capitalists, as well as for the authorities. But it developed to become more independent and focused on reasoning, knowledge and science (Habermas 1991, 15-25). But as we saw in chapter 2.1, Habermas claimed that the media caused the decline of the public sphere in the 20th century as it became refeudalized and commercialized (1991, 158-162).

Another academic whose ideas have shaped the idea of a deliberative democracy is John Dewey. Dewey, an American psychologist and philosopher, was an advocate for democracy who considered schools and civil society to be fundamental. In his text *The Eclipse of the Public* (2003), Dewey describes the American democracy as a collection of local communities with town-meeting practices and ideas, brought together in a national state by the use of technology. Democracy, according to Dewey, was inevitable with the invention of technologies such as the printing press, the telegraph, the railroad and mass manufacturing, as well as the concentration of the population in urban centers. He believed that democracy calls for criticism but described this criticism as “querulousness and spleen” and as being without all-or-none situations. Participatory democracy was the form of society that Dewey believed would best enable all people to lead long, healthy and happy lives (Benson, Harkavy, and Puckett 2007, xii). Dewey asks the question “What is the public?” when considering political apathy in the United States. He problematizes the fact that voters tend to vote against candidates and issues they do not like, rather than voting for candidates and issues they agree with. Dewey considers the ideal public and democracy as a community, arguing the importance of knowledge, insight and communication (Dewey 2003).

Because comment sections serve as a forum for public debate it is tempting to see comment sections as serving an important function in a deliberative democracy. Habermas writes of public debates, open to all, and Dewey emphasizes participation and community, and one can imagine comment sections fulfilling these ideals in one way or another. This does, however, appear to not be the case. While productive debates can certainly take place in comment sections, comments have been found to have little deliberative value. Researchers have found that Internet users do not embrace opinion diversity and that they provide argumentation of little deliberative value (Edgarly et al. 2009). Habermas himself has not been very positive about the internet, calling computer-mediated communication parasitical because internet-based communities have fragmented the public (Geiger 2009, 2). Comment sections also do not fit well with the idea of coming to a decision based on superior argumentation. Whereas public debates among politicians conclude with an election or political decision-making, debates in a comment section have no goal and no end. Commenters do not end a debate in a comment section with a conclusion; there is no vote about who won the debate or an external judge, and commenters rarely reach a shared understanding about an issue. Comment section debates only end when the commenters have made their points and moved on.

Comment sections having a weak standing in terms of a deliberative view of democracy may help explain why many have a negative view of them. If rational-critical debates, reflection on one's own opinions and judgment, community, and the production of unanimous preferences are considered democratically valuable, then comment sections simply do not meet the standards set by society.

Despite the shortcomings of deliberative democracy as a framework for thinking about comment sections, in recent years deliberative democracy has been updated with new thoughts. One of the problems with most research on deliberative democracy, according to Mansbridge et al., is that it has been focused on single, one-time instances of deliberation or as a continuing series of episodes with the same group or the same type of institution, with no focus on the interdependence of such episodes within a larger system. The systemic approach to deliberative democracy, however, recognizes the complexity of democratic entities and examines their interaction in the system as a whole (Mansbridge et al. 2012, 1-2). A deliberative system is, according to Mansbridge et al., "a set of distinguishable, differentiated, but to some degree interdependent parts, often with distributed functions and a division of labour, connected in such a way as to form a complex whole" (2012, 4). A deliberative system has three functions: epistemic, ethical and democratic functions. The epistemic function is to produce preferences, opinions and decisions that are informed by facts and logic, that are the outcome of substantive and meaningful consideration of relevant reasons. Ethical functions are to promote mutual respect, and that citizens should be treated as autonomous agents who take part in the governance of their society. Ethical functions are not limited to mutual respect, but this is brought up by Mansbridge et al. as the primary ethical function. Finally, democratic function is the inclusion of multiple voices, interests, concerns and claims on the basis of feasible equality (2012, 11-12). The systemic approach to deliberative democracy presented by Mansbridge et al. might be more useful when studying comment sections because it provides a more complex framework for looking at the role of comment sections in democratic societies. However, it too focuses on decisions, making its relevancy to comment sections unclear.

The focus on argumentation and reasoning found in deliberative democracy can also be found in discursive theory. Both theories draws on Habermas and focuses on rational discussion and deliberation (Ferree et al. 2002; Strömbäck 2005). Habermas is a central thinker within this discursive theory, as well as central aspects such as the focus on the argument rather than the person making the argument, disregard of status, and rational subjects who's claims may only be accepted if it is supported by valid arguments. As with most democratic theories, discursive theory emphasizes civility and mutual respect, and deliberative theory focuses on trust, integrity and tolerance.

An interesting aspect of discursive theory is that it is based on the assumption that all participants are part of the same moral community and that they share the same basic values (Ferree et al. 2002, 303). People whose opinions are outside of the boundaries of commonly accepted values may not deserve the same kind of respect for their opinions as does others. While this makes sense on an academic level when thinking of racist opinions for example, making the distinctions between commonly accepted values are not always a simple exercise. In some cases, there are no commonly accepted values, but large groups of people with differing values that they both argue are commonly accepted. In a polarized political climate, such as when a country is divided between two major political parties who both claim to hold the "true" values of their nation, discursive theory only makes sense within each group. Comment sections are one of the places where these two groups meet, without such commonly accepted values.

In recent years, new theories of democracy have emerged that focus more on participation and see democracy as a bottom-up process where citizen participation and engagement are highly valued. These theories include agonistic democracy, participatory liberal theory, and participatory democracy. These theories are related to the ideologies that drove the emergence of participatory journalism and indymedia. With the World Wide Web gaining popularity it became possible for anyone to express themselves online. With roots in left-wing social movement activity in 1999, indymedia challenged the professionalism of established journalism. This form of journalism was characterized by a strong political agenda, participatory citizen journalism and a decentralized and localized structure. Editorial policies were replaced with open publishing and the borders between the publisher and the audience became less clear (C.W. Anderson 2012, 82-83). Within participatory liberal theory, the media should represent all interests in society, transform individuals into engaged citizens, and encourage empowerment. It is interesting that the Indymedia movement, which involves participatory publishing, began at the same time as blogging evolved to become more mainstream. Blogging sites like Pitas and Blogger were created in 1999, and in the following years, blogging gained mainstream popularity (Rettberg 2014, 9-14). While bloggers are not professional journalists, their publications can have journalistic content. According to Rettberg, bloggers intersect with journalism in three ways: 1) Bloggers can provide first-hand accounts from events around the world, 2) bloggers can tell stories that are ignored by professional journalists, and 3) bloggers can publish stories about specific topics that are covered in mainstream media (2014, 92-93).

Anderson argues that Indymedia led to the reemerging of agonistic democracy. This form of democracy is described as a vibrant clash of democratic political positions. It is thought that too much emphasis on consensus and a fear of confrontation leads to apathy. Disagreement is unavoidable in a democracy, and should not be feared, according to proponents of agonistic democracy (C.W. Anderson 2012, 91-92). Agonistic democracy involves the contesting of basic principles, conflict, attention to the informal operations of power, and respect for differences (Wingenbach 2011, xi). Indymedia is focused on open participatory publishing, and comment sections are by their very nature a form of open participation. Debates in comment sections have, as noted earlier, been described as having little deliberative value, and they are without a goal or an end. Debates in comment sections are debates for the sake of debating in which conflict is the norm – conflict that should not be feared, according to agonistic democratic thought. However, there is one part of agonistic democracy that may be lacking in comment section: respect for differences. Comment sections are, as we have seen, places with antagonistic discussions where participants are not above using slurs and acting in an uncivil manner against other commenters.

Agonistic democracy seems to be related to ideas within participatory liberal theory. According to Ferree et al. (2002), this democratic theory focuses on maximizing the participation of citizens in the public decisions that affect their lives. This theory has roots in Rousseau's preference for direct democracy, and proponents have a distrust of institutional barriers that makes participation indirect and difficult. Citizen participation, according to Ferree et al. should be an ongoing process by grassroot actors. While the focus of participatory liberal theory is not as antagonistic as agonistic democracy, both theories entrust and encourage a broader population to have the capacity to participate in democratic processes. Another participatory democratic theory is, as the name suggests, participatory democracy. This theory, according to Strömbäck (2005) states that people are expected to be engaged in civic and public life, and participate in community activities.

Within participatory liberal theory, individuals should be empowered and are encouraged to be engaged. One can question how empowering comment sections are, but commenting on public articles could certainly be a forum for engaging individuals. However, one could argue that toxicity in comment sections does not encourage engagement. Finally, comment sections can be argued to not fulfill the requirements of participatory democracy either, as this theory states that citizens should not distrust each other (Strömbäck 2005). Distrust can be seen in comment sections as accusations of factual errors, ulterior motives or trolling.

Mouffe (2022, 1-3) argues that following the 2008 economic crisis there was a populist movement in response to thirty years of neoliberal hegemony. Populism is described by Mouffe as "a discursive strategy of constructing a political frontier dividing society into two camps and calling for the mobilization of the ‘underdog’ against ‘those in power’ (Mouffe 2018, 10-11).[2] Mouffe claims that “these transformations have led to a situation referred to as ‘post-democracy’”, because equality and popular sovereignty has been eroded. Central to Mouffe’s argumentation is that there is a consensus between the center-left and the center-right that there is no alternative to “neoliberal globalization”. This means that traditional parties no longer offer any opportunity do decide on real alternatives in an election, and Mouffe claims that this has led to one of the fundamental pillars of democracy being undermined: popular sovereignty. Furthermore, Mouffe argues that financialization of the economy and the 2008 economic crisis has eroded the other pillar to democracy: the defense of equality.

Mouffe is a political actor who describes herself as having political objectives (Mouffe 2018, 10). Despite this, her ideas have relevancy to the later discussion about the role of comment sections in a democratic society. The post-democracy that Mouffe describes has been responded to by populist, anti-establishment movements (Mouffe 2022, 3). One could argue that the rise of indymedia is another form of anti-establishment movement. If one were to see the traditional news media as part of the establishment, then digital news sites created by non-professional journalists could be seen as anti-establishment. The same could be true for blogs, and perhaps even comment sections.

[2] This definition is by Laclau (2005), but it’s the definition that Mouffe chooses to use.

5. Researching comment sections

Comment sections provide researchers with a lot of data. Millions of comments across thousands of online newspapers are publicly available, creating a possibility for researchers to perform qualitative and quantitative studies on real-world examples of online behavior. As opposed to studies performed in laboratory conditions, studies that sample real-world comments from newspaper comment sections are based on a dataset that represents how people truly behave online. But studying comment sections is not unproblematic. There are several methodological and ethical issues to be aware of.

5.1. Methodological challenges when studying comment sections

For my research projects, I investigated comments sampled from three U.S. news sites. When sampling comments from a newspaper's comment section, timing is important. Comment sections are not static, but an evolving and changing forum where users can continually add to the data. I would therefore recommend that any sampling should be done when enough time has passed since the publication of the article that one can expect that anyone who wants to comment have already done so, to ensure that all comments are represented in the data. In preparation for my master thesis I investigated the longevity of commenting on twelve articles from the Norwegian newspaper VG, and found that commenting would occur on an article for up to four days after the article's publication (M.A. Knustad 2018, 17). While this is not a representative finding that can be generalized to all comment sections, it indicates at the very least that several days can go by before commenters have stopped commenting on an article. And even then, there is nothing to stop a commenter from writing a comment month, or even years after the publication of an article. When sampling comments for this thesis, the chosen articles were published between one month and thirteen months prior to sampling. Though I cannot guarantee that comments have not been made on these articles after they were sampled, I believe it's reasonable to expect that very few, if any, comments would be made that late after the publication of an article.

While sampling comments long after an articles publication is a good way to ensure that no new comments are made after sampling, another problem arises. The researcher will not be able to see comments that, for whatever reason, have been deleted by moderators or the commenters themselves. Commenters may delete comments because they regret writing them, and moderators may delete comments that do not comply with a newspaper's rules of conduct or are illegal due to their content. Such deletions may occur soon after a comment was written, and the only way to also sample these comments would be to perform the sampling in real time – assuming moderators would delete a comment after publication as opposed to approving or deleting comments before being published.

A researcher may choose to study an article that was just published and continually collect comments as they were written. However, this would be a very time-consuming methodology, as a researcher would have to constantly check for new comments, potentially for days. It would also mean that the selection of articles to be studied would have to be done based on which articles are being published during the sampling period, making it difficult to choose articles based on a particular topic that the researcher might be interested in. And, of course, this method would be impossible for research projects with a more historical focus where one might want to study comments over time.

In my own research, I chose to use a strategy for sampling called constructed week sampling, which will be explained in more detail later (Chapter 5.3.). Because this methodology involves collecting comments from a longer time period, it would have been very impractical for me to collect comments in real-time. Had I done so I would have had to spend one year on sampling alone. In my own research, I am aware of the possibility that comments may have been deleted before I had sampled them. However, seeing no reasonable way to avoid this, I must simply acknowledge that this possibility exists and move on. In the end, I can say that I have studied comments that have been published on comment sections without being deleted, and not comments that have been written on comment sections.

5.2. Ethical challenges when studying comment sections

5.2.1. Informed consent in big data projects

When sampling comments for my study, comments were sampled without informed consent from the commenters. As described below, this is because consent would be difficult to acquire and because comments are published in a public space. My research on comments involved an observational research design, meaning that as a researcher I could only observe behavior in comment sections, without any interventions other than to record, classify, count and analyze the results (Porta 2014, 204). When studying comment sections, the participants being observed are the commenters. They are, however, not willing participants, and have not given informed consent to be a part of any research study. Informed consent is, according to Faden and Beauchamp (1986, 3-4), rooted in the fields of law and moral philosophy. The law is primarily focused on clinical contexts and has the pragmatic goal of reducing risk. Moral philosophy, on the other hand, is primarily focused on the respect of autonomy and an individual’s right to make an autonomous choice.

When making ethical considerations, it is important to consider the potential risks and benefits of the research and its methodology. The potential risk of my research was the exposure of commenters. These individuals were sharing their views on sensitive political topics, something that may carry personal risks if they are exposed. If an academic article or thesis exposes a person's political views, that exposure would be more permanent than the original comment itself, as comments can be deleted. Therefore, researchers should implement appropriate measures to protect commenters' identities, and limit exposure. How extensive those measures should be is in part affected by the benefit of the research being performed. The potential risks to the commenters should be weighed against the beneficence of the research (Pieper and Thomson 2016). While I acknowledge that there is a potential risk to the commenters whose comments are used in my research, I would argue that this risk is low. The commenters have willingly shared their opinion on a public platform and steps have been taken to ensure that their exposure is as low as possible. Therefore, I would argue that the potential risk to the commenters is outweighed by the potential benefits of the research. By examining the democratic role of comment sections, this research could potentially help our understanding of how to best facilitate a democratically valuable public debate. This could potentially benefit news sites, commenters and society. This would, however, be difficult to accomplish when sampling large amounts of data from newspaper comment sections. In its ethical guidelines, the Association of Internet Researchers (AOIR) expresses concerns surrounding the problem of retrieving consent when doing research on big data projects where the data is retrieved from semi-public spaces, and especially when such data may contain identifiable or sensitive information (franzke et al. 2020, 10). While I would argue that my own research is done in a fully public space, I must acknowledge that the full names of the commenters, and perhaps even pseudonyms, can be identifiable information.

In my own research, I sampled 3851 comments written by 2401 commenters. Retrieving permission from this many people would not be an easy task. Firstly, many of them use pseudonyms. One could argue that there are no privacy concerns when people are using pseudonyms, but this view would not take into account possible ways that a pseudonym could potentially be linked to a real identity on other platforms. As an example of this, consider how the Silk Road founder Ross Ulbricht was identified by the FBI because Ulbricht had used a pseudonym across different platforms, one of which contained a real-life e-mail address (Hume 2013). While this may be an extreme example, it shows how the use of pseudonyms across platforms can allow someone to uncover the identity of an individual. Despite this, I would argue that it is unreasonable to expect that a researcher should use such investigatory efforts to identify an individual in order to retrieve informed consent. Uncovering who the anonymous commenters are is not practically possible considering how difficult it would be to uncover their real identities. The non-anonymous commenters, however, are easier to identify. But I would argue that retrieving permission to use their comments in research is very problematic. For each commenter, the researcher would have to essentially track down the contact information of the commenters based on their names alone. Again, this would require investigatory efforts by the researcher that I think is unreasonable to expect.

Alternatively, one might use the comment section itself to contact the commenters, or at the very least write a comment explaining that the comments on this article will be used in research and providing an opportunity to opt out of participation. This, I would argue, is a problematic solution for three reasons. Firstly, it is a very pour way of communication, especially on comment sections of articles written months or years ago. I would expect the response rate to be poor at best. Secondly, this could be considered a misuse of the comment section itself, and a possible violation of the news site's rules of conduct. Moderators may not consider requests for participation in research and attempts to establish contact with the commenters as a proper use of the comment section. It is certainly not what comment sections are meant to be used for. Finally, I would argue that it is problematic for a researcher to, in essence, participate actively within the group of participants, which is what the researcher would be doing by leaving any comments in the comment sections that are used to sample data from. Participation by researchers may be warranted in experimental research designs, but not observational ones, where the study should not involve any intervention on the part of the researcher (Porta 2014, 204).

In my own research I did not find it practically possible to get consent from any commenters, and with permission from the Norwegian Centre for Research Data I collected comments without any form of consent or notifications about my research project. It is my own opinion that commenters fall within a special category of research participants, as they have volunteered the information used in research publicly. According to the Norwegian National Research Ethics Committees, consent may not be necessary when studying public expressions (*Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora* 2021). I would argue that commenters publish their comments in a public forum. In fact, the comment sections of national newspapers are just about the most public place people can express themselves online. While I'm sure the commenters never intended for or considered the possibility that their comments could be used in research, the same can be said for many forms of expression that may be studied by researchers. Where does one draw the line between which forms of public expression can or cannot be studied at will? There is a wide range of such expressions, from public Facebook profiles to forum posts, blogs, opinion pieces in newspapers and articles. While I have no clear answer to where one would draw this line, I would argue that newspaper comment sections are clearly public spaces, and that commenters should be aware of this.

That being said, when dealing with ethical questions in research it is usually a good idea to be cautious with potentially identifiable information. Without informed consent, identifiable information should be handled with care. According to the Association of Internet Researchers, one way to mitigate risk against research subjects that have not been given the chance to provide informed consent, is to delete names and other identifiable information when storing and processing the data (franzke et al. 2020, 10). In my own sampling process, I continuously deleted both names and pseudonyms from the dataset and replaced them with numeric identifiers, ensuring that there was no identifiable information in my data.

5.2.2 Presentation of the research

Privacy concerns are not only important when sampling data. Even when the data is anonymized during sampling, measures must be implemented during storage and the presentation of the results. The Association of Internet Researchers emphasizes the importance of considering how data is being stored and presented, and what measures are taken to secure data (franzke et al. 2020, 19). In terms of storage, the anonymized data used in my research was stored on a university server space that only I had access to. In addition to this I implemented several security features that I will not go into in any detail.

Presentation of the data is another area of ethical concern. When reporting on findings, researchers should only report what is necessary to make their points. That is why I, in my own thesis and articles, will not mention specific articles that comments have been sampled from, at what time, or the name or pseudonym of the commenters. All of this is information that could help to identify commenters, and it is not information that is necessary in the presentation of my results.

According to Markham and Buchanon, even anonymized datasets can contain enough personal information for an individual to be identifiable (2012). In other words, the very content of a comment could potentially lead to a commenter being identified. Even if it contains no personal information, it may be possible through search engines to find a comment in its context, thereby risking the commenter being identified. I have tried using this method myself on comments from the dataset and found that comments do not show up in search results. Comment sections are usually embedded on an article through an iframe that scripts, such as the web crawlers that search engines use for indexing websites, have difficulties reading. But one of the issues raised by The Association of Internet Researchers is if future technologies make it possible to strip personally identifiable information from data sets (franzke et al. 2020, 19). Due to changing technologies, search engines may index comment sections in the future, and steps taken to ensure the privacy of research participants should be future proofed as much as possible.

As mentioned earlier, getting participants' informed consent is problematic when sampling data from the internet. One of the solutions to this problem, as suggested by The Association of Internet Researchers is to reserve the acquisition of informed consent to the dissemination stage of a research project (franzke et al. 2020, 10-11). This means that participants will only be contacted if their data is going to be presented in the research, such as when using comments in a research article to exemplify certain traits or behaviors. Because so few participants have to be contacted, getting informed consent becomes a much more manageable exercise. In my own research, however, this strategy would be challenging. As mentioned earlier, I continuously deleted identifiable information during my sampling process. This means that there was no way for me to easily identify the author of a comment. And even if I had managed to find the name of the author, the same difficulties of contacting the author as described above would still apply.

In my own reporting I will not, in accordance with the requirements of the Norwegian Centre for Research Data, quote any comments in my thesis or any academic articles. Even anonymized comments, presented out of context, could be thought to contain personally identifiable data. Future technologies may make it easy to search for and identify the writers of comments presented in academic articles and theses, and in my own research every precaution will be taken to assure that potentially identifiable information is not accessible. Not being able to show examples of comments can be limiting, especially in qualitative research. In cases where examples are warranted, I have instead used paraphrasing and descriptions to illustrate important concepts regarding my own findings.

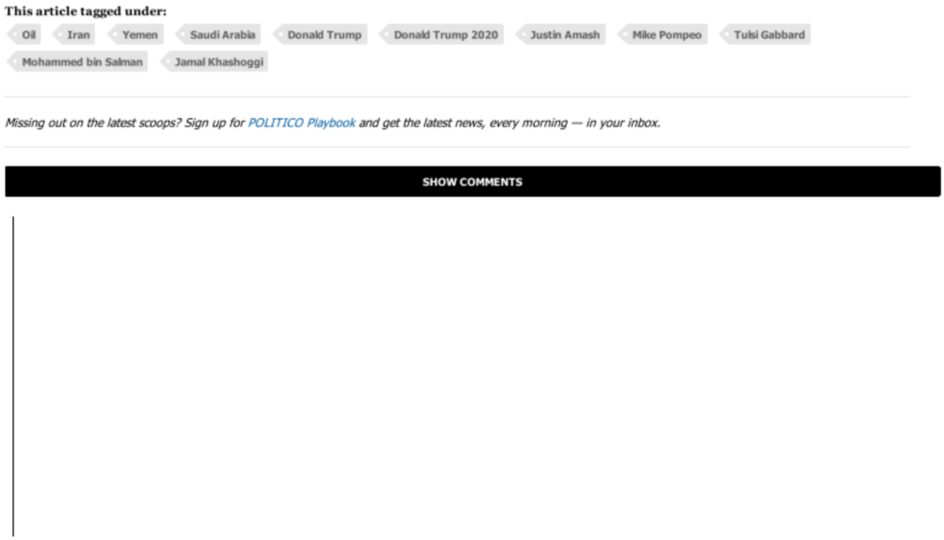
5.3. Sampling

The research for this thesis involved three research projects looking specifically at comment sections, with varying methodologies chosen based on which methodology would best answer the individual research questions. These projects used the same sample of 3851 comments from Politico, The Washington Post and the New York Times.

There are several reasons why American news sites were studied. To do this research, I had to choose comments written in a language that I understand and within a cultural context I'm familiar with. Choosing comments in a language I do not understand, would result in the need for translation services. And sampling comments from a culture in which I won't understand the context of the comments would make it difficult to perform a proper qualitative analysis of data that might consist of culturally specific terms, idioms, and references. I speak English and Norwegian, therefore I decided to sample data from Norway or an English-speaking country. Norway and the U.S. are the two countries where I would most likely understand the cultural context of the comments, as I have lived in both and have extensive knowledge of the culture and politics of both countries. Therefore, I rejected sampling comments from any other country. In the end I decided to sample comments from American news sites because the U.S. provided a broader range of news sites and comment section technologies. I found this to be important, especially at the beginning of the research project, because it opened more opportunities for research. While this factor became less important later on, I think it was a reasonable choice to make because it would be better to have access to more news sites and platforms and not need it, than the other way around.

Sampling was done using constructed week sampling. Using this method, two constructed weeks from February of 2018 to February of 2019 for each newspaper being studied were created. This involved selecting two random Mondays, two random Tuesdays, etc., during the specified timeframe. This method of sampling is recommended for studying daily newspapers because it creates a randomly selected issue for each day of the week. Two constructed weeks have been found to be sufficient for representing a year's content (Riffe, Lacy, and Fico 2014, 85-86). 3851 comments were collected and stored in a database using this method.

Finding an efficient method for collecting comments that also satisfies the requirements for user data protection by the Norwegian Centre for Research Data (NSD) has proven to be difficult. The ideal method for collecting comments would involve some sort of automated process, such as a bot scraping comments sections of various online news sites. The first obstacle with creating such a bot is that most comment sections are found in iframes. Iframes are HTML-elements used to embed another document within a web page. This mean that an article's comment section is not found in the same HTML-document as the article but exists in a separate document that is embedded underneath the article. For the human reader of a comment sections, this separation and embedding has no consequence, as the embedded comment sections appears on the screen like any other element. But computer programs can have difficulties reading content in iframes. As an example, consider NCapture, a browser extension used to capture web pages and download them to NVivo, a licensed software used for qualitative research. In NVivo, a web page can be seen as a PDF-file. But as shown in Figure 1, the NCapture browser extension has failed to read the comments in the comment sections on Politico.com.



This article tagged under:

Missing out on the latest scoops? Sign up for POLITICO Playbook and get the latest news, every morning — in your inbox.

Figure 1. Screenshot from NVivo, showing how NCapture has failed to read the comment section from an article on politico.com.

While it may be possible to find a solution to the technical problem of computer programs not reading content in iframes, there is also an ethical problem that makes any automatic collection of comments problematic. To get my research project approved by NSD it was necessary to, among other things, anonymize all data as soon as possible. Finding a technical solution to identify and anonymize names is very difficult. While some names are marked in the HTML with certain class-names, commenters may choose to write the name of other commenters in plain text. Therefore, it became necessary to collect the comments on a one-by-one basis, where each sampled comment was carefully read and copied one by one. The comments were copied into a purpose-built script along with metadata, such as whether the commenter was anonymous, response level and date of publication. This information was stored in a database on a university computer. In a second database I stored the names of each commenter. These names were matched with numeric identifiers that were then stored in the first database with the comments themselves. This strategy allowed me to separate different commenters in the data, without using anyone's real names. To further protect the privacy of the commenters, the data in the second database was continuously deleted. This was done after I had finished collecting comments from each newspaper. At the end of the research project all data was deleted in accordance with the NSD-requirements.

6. Discussion

6.1 Anonymity and participation

In the 1970’s, some authors saw the emerging internet technologies as a possible way to revitalize democracy and stimulate public debate (Gonçalves 2015). Now, half a century later, we know that internet technologies can cause unique challenges. Online toxic behavior has been a field of study since the 1990’s (Suler and Philips 1998; Jessup, Connolly, and Galegher 1990; Philips 1996; Wang and Yan 1996), and many researchers have pointed out toxic behavior in comment sections (Lapidot-Lefler and Barak 2012; Rowe 2014; Stroud, Muddiman, and Scacco 2016).

In the literature review of this thesis, several democratic theories were considered, and comment sections were found to be problematic within most of them. Most democratic theories I have studied emphasize in one way or another the importance of mutual respect when disagreeing on a topic. There is enough scientific and anecdotal evidence to suggest that disagreements in comment sections are not met with respect. Even when the commenters are not behaving in a toxic manner, discussions tend to be antagonistic in nature.

On the surface, theories like participatory liberal theory seem to be a good fit for comment sections. Participatory liberal theory is centered around citizen participation in public decisions and argues for direct democracy and the minimization of institutional barriers. Comment sections, then, seem to be of great value within this framework. Journalism is traditionally a field of strong institutions, where a few gatekeepers control the flow of information in a one-to-many broadcasting system. Allowing for citizen participation in journalism by facilitating commenting on news articles can be argued to break down some of the institutional barriers of journalism, meaning that the role of comment sections in a democracy is to facilitate freedom of expression and public discourse. However, this assumes that everyone has equal access to the discussions taking place in comment sections. On a technical level they do, but the prevalence of toxicity (Rowe 2014; M. Knustad and Johansson 2021; Gonçalves 2015; Vergeer 2015) may scare some people away from participating in this form of public discussion (Stalsberg 2015; Stroud, Van Duyn, and Peacock 2016; Rossini 2019). Proponents of participatory liberal theory wish to maximize citizen participation, but the prevalence of toxicity in comment sections and the resulting negative impression of them may cause people not to participate due to fear of negative reactions from other commenters. As such, when seen through the framework of participatory liberal theory, the role of comment sections in society is not a positive one because of the prevalence of toxic comments.

Toxicity then, is a major factor when discussing the role of comment sections in a democracy. As we have seen, toxicity is prevalent in comment sections. For the most part, two strategies are employed to combat toxicity: moderation and removing anonymity. Moderation, where newspaper employees or volunteers delete toxic comments, may lead to new toxic comments when those who have had their comments deleted will angrily question the deletion, often mistakenly arguing that their right to freedom of speech have been denied. Commenters who have experienced having their comments deleted tend to be more negative towards moderation than others, and tend to have a non-interventionalist view of moderation, where those who have a poor impression of comment sections have a more interventionalist view of moderation (Løvlie, Ihlebæk, and Larsson 2017). Moderation then, necessary as it may be at times, could further antagonize already toxic commenters and possibly other commenters who agree with the moderated individual's point of view.

The other strategy of removing anonymity is also problematic. Anonymity is an often-cited explanation for online disinhibition (Rowe 2014; Santana 2014; Bae 2016; Barlett, Gentile, and Chew 2016), and there has been a move in recent years to either close comment sections or not allowing anonymous comments (Bilton 2014; Ellis 2015; Wallsten and Tarsi 2015). However, there are many legitimate reasons why someone would want to remain anonymous, including situational motivations that arise when posting things with one's real name makes it possible for two different posts from different sources to be presented in the same search results (Hogan 2013). It has also been argued that the use of pseudonyms make it possible to avoid context collapse (Marwick and Boyd 2010). Furthermore, it is important to note that having multiple identities is a natural part of the human experience, and it could be argued that this should be extended to our online lives (Mead 1934; Suler 2016, 73). Another reason why anonymity may be important in comment sections is that it lowers the bar for participation, allowing more voices to be heard in this public forum. If fear of harassment and toxicity makes people avoid comment sections, as research suggests (Stroud, Van Duyn, and Peacock 2016), then I think it's reasonable to question if allowing people to comment anonymously would counter this effect.

With all these arguments in favor of anonymity it is concerning that an increasing number of online publications require commenters to identify themselves and post using their real names. And it is especially concerning that several such publications use a comment section plugin from Facebook to facilitate real name commenting. This not only makes a Facebook-account a requirement for participation, but it further blurs the lines between different online identities and raises concerns about privacy (Reagle 2015, 8-9).

The counterargument is, of course, that comment sections become more civil without anonymity. This could also lower the bar for participation, as toxic comment sections may scare away people who wish to contribute to the discussion. However, considering the many arguments in favor of anonymity, the positive effect of requiring commenters to identify themselves must be substantial enough to justify the potential negative effects of not allowing commenters to be anonymous. As we saw in the literature review, several studies have shown that anonymity influences civility in comment sections (Rowe 2014; Santana 2014; Lapidot-Lefler and Barak 2012; Barlett, Gentile, and Chew 2016). But as I discuss in my article, *Anonymity and inhibition in newspaper comments* (M. Knustad and Johansson 2021), such studies can have problematic methodologies. In my own study we analyzed comments from The Washington Post and The New York Times, which include both anonymous and non-anonymous commenters. This combination of anonymous and non-anonymous commenters on the same platforms provided a good opportunity to study the effects of anonymity because both the anonymous and non-anonymous comments are retrieved from the same platform, ensuring that any differences between them should be due to anonymity.

The results of the study showed that there was a weak but statistically significant relationship between anonymity and toxic comments.[3] This relationship seemed to be caused by non-anonymous commenters writing fewer toxic comments than what would be expected, meaning that anonymity didn't cause toxicity as such. Instead, not being anonymous caused commenters to be less toxic. However, the observed relationship was weak, meaning that anonymity cannot be used to fully explain toxic disinhibition in comment section. Therefore, other possible explanations for why people behave in an uncivil or impolite manner in comment sections must be considered.

There is clearly a link between anonymity and toxicity, mas demonstrated in multiple studies (Rowe 2014; Santana 2014, 2019; Bae 2016), but I would argue that we should be cautious about making definitive explanatory conclusions based on these studies. In this thesis, I have outlined other possible explanations for toxicity, and we must have a serious discussion about the value of anonymity versus the potential benefits of not allowing anonymous comments. I would argue that not allowing anonymity could devalue the democratic potential of comment sections by raising the bar for participation and contributing to individuals experiencing context collapse. This is a subject that should be further investigated and discussed in relation to the findings of anonymity as a cause for toxicity, and other possible explanations for toxicity should be researched further.

My study of anonymity in comment sections not only show that there is a small relationship between anonymity and toxicity, meaning that removing anonymity may not be a good solution to combat toxicity. It also shows that currently toxicity is to be expected in comment sections regardless of whether comments are anonymous. And if we must expect toxicity in comment sections, then we may also have to accept that comment sections do not serve a positive role in society from the perspective of participatory liberal theory, because toxicity could raise the bar for participation – which is the main focus of participatory liberal theory.

[3] Toxic comments are all comments that were coded as being either uncivil or impolite.

6.2. The problem with judging comments by the standards of deliberative democracy

But what about deliberative democracy? When comment sections are criticized in research, it is often done within a deliberative framework (Løvlie 2018, 3). I would argue that deliberative democracy or discursive theory are not very useful as a framework for analyzing comment sections. Deliberative democracy involves open, accessible and critical debates with thoughtful arguments. The goal is to reach a conclusion based on the superior argumentation. There is no such goal in comment sections, however. In comment sections, there are no final votes to end a discussion and determine which viewpoint is the "winner" based on superior argumentation. The very nature of comment sections is that the debate is open-ended, with no endpoint where one can say that the debate is over. The end of a debate in comment sections is a passive end; the point at which no one is commenting anymore, as opposed to an active end with a definite endpoint. In addition to this, debates in comment sections can be very broad. A single article can result in countless discussion threads on varying subtopics, further complicating the thought of comment sections being used to reach a consensus based on argumentation. Comment sections arguably do not live up to the standards of classical deliberative democratic theory. However, considering the very nature of comment sections, I would argue that it is unfair to hold them to such a standard. The central aspect of deliberative theory - reaching a conclusion based on the superior argument - is in direct conflict with the very nature of comment sections, where multiple debates can take place at once without any of them working towards a conclusion.

The systemic approach to deliberative democracy, however, might be a useful framework because it recognizes the complexity of democratic entities and examines their interaction in a system as a whole (Mansbridge et al. 2012, 1-2). This theory allows us to view democracy across three dimensions; the epistemic, ethical, and democratic functions. An ideal democratic institution should produce decisions that are informed by facts and logic (epistemic function), promote mutual respect (ethical function), and include multiple voices, interests and concerns (democratic function). But again, we see an emphasis on decisions, making it unclear if this theory can be used to analyze comment sections. As mentioned above, there is no final goal or decision in comment sections.

6.3 Trolling and not respecting each other's differences

Comment sections could be argued to have an important role in the framework of agonistic democracy. As opposed to the standards of deliberative democracy, where the goal is to reach a conclusion based on the superior argument, proponents of agonistic democracy places value on the discussion itself rather than reaching a consensus. Confrontation or conflict should not be feared, as it is the disagreement and the ability to voice one's opinions that is democratically valuable. However, we must again consider the implications of toxicity in comment sections. Agonistic democracy involves having respect for differences, something that is often lacking in comment sections that are polarizing and have problems with toxicity (Vergeer 2015; A.A. Anderson et al. 2018; A.A. Anderson et al. 2014). Again, we see that toxicity is problematic, and that because of toxic comments comment sections may not have a positive role in society. Furthermore, my research into accusations of trolling further showed how participants in comment sections can show disrespect for each other's differences – specifically differences in political opinions.

Researching trolling is quite difficult because trolling behavior must be identified by the researcher. The label of *troll* can mistakenly be attributed to genuine comments, that for whatever reason are shocking and disruptive. In my article, *Get lost,* troll (M. Knustad 2020), I did not focus my research directly on trolls and trolling behavior. Instead, I studied how accusations of trolling affected the debate in comment sections. Trolls can affect the debates not just by their presence and their actions, but the concept of trolling and the knowledge that trolls exist, can also affect the debates in multiple ways, as my study showed.

With the general public becoming more aware of trolling (Dimock 2019), one could hope that participants in comment sections would be able to identify trolls and act accordingly (T. Graham and Wright 2015). I hypothesized that as the public become more aware of trolls, bots and foreign influence in social media, accusations of trolling could be used as a rhetorical tool to delegitimize and discredit an opposing argument. If this were the case, it would be problematic for the deliberative value of comment sections as a forum for public debate. Furthermore, it is important to consider how the users of a comment sections react to trolling as a group. Accusations of trolling, whether they are true or not, should at the very least elicit concern from commenters. My study aimed to uncover both how the group as a whole reacted to accusations of trolling and how individuals accused of trolling reacted.

In this study I made four conclusions about how accusations of trolling affect the debate in the studied comment sections:

1) Most accusations of trolling were made by left-wing commenters and directed towards right-wing commenters, suggesting a clear political difference between those accusing and those being accused of trolling.

2) It was common for accusations of trolling to be motivated by political differences, rather than the rhetoric used by the accused. While there were accusations made towards commenters writing divisive, conspiratorial or vulgar comments, many similar comments did not lead to such accusations. In addition, about half the accused commenters wrote argumentative or informative comments simply expressing their opinions.

3) Most of the commenters accusing someone of trolling would either challenge the accused troll's arguments, make fun of the troll, or warn other commenters about the presence of a troll.

4) Accusations of trolling were rarely responded to by the accused person or other commenters, suggesting that such accusations are mostly ignored by commenters engaged in a debate.

My article shows that instead of comment sections being able to self-regulate in the presence of perceived trolls, warnings about trolls were mostly ignored by other users who continued to engage with the suspected troll. This means that real trolls would likely be taken seriously by their fellow commenters, allowing trolls to perform their disruptive behavior unchecked. Furthermore, accusations of trolling were found to be used as a rhetorical device to delegitimize opposing arguments. These results further illustrate how the role of comment sections, especially in terms of their value as a platform for public discourse, is unclear. Within agonistic democracy, confrontation should not be feared as it is the disagreement and the ability to voice one's opinions that is democratically valuable. I would argue that a platform that is susceptible to trolling and where opposing arguments are delegitimized without being considered or argued against is not a valuable contribution to public discourse.

6.4. Commenters as gatewatchers

Perhaps the role of comment sections is not to be a positive contribution to deliberation and participation, meaning that all the democratic theories I have discussed are irrelevant in this context. Perhaps the role of comment sections is to be a counterweight to the establishment. The post-democracy that Mouffe describes (2022, 1-3) is relevant to the discussion of the role of comment sections. Mouffe argues that the post-democracy that has emerged because of globalization and neoliberalism has been responded to by populist, anti-establishment movements.

For there to be an anti-establishment, there must first be an establishment. In the context of comment sections, the establishment are the newspapers posting the articles that are commented upon. Comments, such as those being studied in this research project, could be described as non-professional publications attached to professional articles. And it is perhaps within the professional vs. non-professional or the establishment vs. anti-establishment dimensions that comment sections become the most relevant. Comments on articles in traditional media could be seen as an anti-establishment response to the publications of the establishment. Comments, that have been described as "Freedom of speech from the depths of the people" (Korslien 2014), represent a portion of the public with little hope of expressing themselves through professional articles. Comment sections provide them with a platform to express themselves in the same space as the “establishment”. And one result of this meeting between professionals and non-professionals, is critique of the traditional establishment media.

The only ones who can truly answer the question of whether or not critical comments are of value to the news organizations are the news organizations themselves. And it is reasonable to suspect that different publications have different views on this. However, assuming that journalists and editors are as human and fallible as the rest of us, the very existence of a direct line of communication on an article that readers can use for criticism ensures that bad journalism will have consequences in the form of critical comments. Comment sections are a unique form of public expression because they allow for public criticism of the media organizations hosting the comment sections. As discussed earlier, researchers and academics have argued that comment sections are a form of participatory journalism, and journalists have reported that comment sections can have a positive impact on their work (Løvlie 2018; T. Graham and Wright 2015). If one were to consider a newspaper as part of the democratic debate, then such criticism should be welcomed, as it could produce higher quality journalism based on fact and logic.

While I would agree that the possibility of journalists receiving criticism in an article’s comment section may influence their work, I felt it necessary to investigate such criticism more closely to better determine if it is valuable. In the article *Critique of the media in newspaper comment sections*, I investigated how users of comment sections criticize the media. I found a relatively small number of comments critical of the media, though I acknowledge that my sampling method may not have caught all instances of media criticism. Only 1.79% of the sampled comments contained criticism of the media, showing that comment sections are not being widely used as a form of media criticism - constructive or not. Constructive criticism is only valuable when it reaches the persons being criticized. And it is unlikely, and unreasonable to expect, that journalists and editors take the time to read as many comments as they would have to in order to find constructive criticism from their readers. One study reported that while some journalists argued that the comments made them reflect on their work, they only read the first fifty comments on their articles (T. Graham and Wright 2015). That is not to say that commenters cannot serve as gate watchers. If an article contains enough errors, the share number of commenters pointing this out would presumably be enough for the journalist or editor to at least double check their work. These would be rare occasions, though. Journalists, especially those working within mainstream media, are expected to uphold a certain standard. And I would argue that obvious or massive errors would be pointed out by other journalists, fact checkers or authors of opinion pieces, meaning that commenters are not a requirement for such criticism to take place.

In my article I present a two-dimensional coding scheme that divide comments by type (quality, integrity, and focus) and target (journalists, news organizations, and the media). I argue that this coding scheme is a scientific contribution because it provides an opportunity to study media criticism in a multi-dimensional way that can provide more insight than a coding scheme with only one dimension. For example, a general media skepticism was found in comments criticizing the media (target-dimension), and especially those criticizing the integrity of the media (type-dimension). These comments are, if we consider the media to be the establishment, truly anti-establishment, in that they showed little trust in the media as a whole and lumped the media in with other societal institutions. This sort of conspiratorial criticism may not be very useful for the media, however. Other critical comments, however, were considered to be more constructive – meaning that the criticism contained specific complaints and/or suggested solutions. Comments targeting journalist or news organizations were more specific and often contained suggestions for how to improve an article, especially those criticizing journalistic quality. This finding shows how critical comments can be useful for improving the journalistic work of a news organization, and hopefully a coding scheme such as the one developed for this research can be useful when creating systems to organize and react to comments. After all, constructive criticism doesn’t matter much if it doesn’t reach the right individuals and isn’t acted upon.

Whether or not critical comments are useful, depends on how the recipients handle the criticism – if they handle it at all and do not ignore it. Considering the massive amounts of comments a news site receives at any given day, it is difficult to imagine that each comment is given the appropriate attention to determine if it is useful in any way. Therefore, the usefulness of comment sections as a forum for criticism of the media, is more dependent on the media's ability to receive them than the comments themselves.

Of course, there are other considerations than how the news organizations themselves receive criticism. The criticism of the media published in comment sections are public and can be read by anyone. Research has shown that comment sections can have an effect on a reader's perception of public opinion, and change the reader's personal opinions (Toepfl and Piwoni 2015, 467). In addition, aggressive comments are more likely to change a readers opinion about the source text (the article), more so than neutral or positive comments, and that such comments can be polarizing (A.A. Anderson et al. 2018; A.A. Anderson et al. 2014). This means that there exists a chance that criticism of the media published in a comment section could have an effect on readers' opinion on the trustworthiness of the media. Media criticism in comment sections could be thought to add to the current state of media distrust and political division. However, one could also argue that the affordances of comment sections make it easier to warn other readers about mistakes or problematic content in an article. The news media is a broad industry with various actors, some of whom do not adhere to a strict journalistic standards or unbiased reporting. Comment sections allow knowledgeable readers to point out such issues. And it is also possible for commenters to share their own experiences on a certain issue that may not reflect the contents of the article, thereby adding knowledge to the issue being reported on.

It could be argued that journalists by simply knowing that they may be criticized by scrutinizing comments might take steps to avoid such criticism. Being more critical and thoughtful of their own work may be one way of doing this. While this won't stop commenters from making unreasonable or toxic criticisms, the feeling of being not only watched, but held accountable by the readers, could be thought to have a positive effect on how journalists work. One must, however, also consider if such a feeling is healthy for journalists who, presumably, would try to do a good job regardless of the existence of comment sections. There is also the problem of non-constructive, unreasonable and sometimes harassing criticism.

In conclusion, comment sections may serve a role as a sort of anti-establishment platform on established media news sites, in which non-professionals may critique the work done by professional journalists. News media are an important part of a democratic society, and any criticism that can help improve the work done by journalists and editors should be welcomed. Any form of audience feedback could be valuable. Though, as my research shows, different types of criticism directed at different targets may be more or less constructive.

6.5. The value of having commenters engaging with news sites

So far, we've determined that comment sections are problematic when seen through various democratic frameworks, that toxicity is a problem that cannot be fixed by removing anonymity (if it can be fixed at all), that comment sections are susceptible to trolling, and that accusations of trolling are used as a rhetorical device to shut down opposing arguments. Despite this, I've come to the conclusion that comment sections could be seen as playing a role as an anti-establishment platform on professional news sites, and that they have the potential to serve as a forum for constructive criticism of the media. Furthermore, we must consider what happens to the commenters if comment sections are closed. According to uses and gratifications theory, individuals seek out the media that fulfill their needs and leads to gratification (Whiting and Williams 2013). And for some users, that need goes beyond just getting updated on the news. The mere existence and popularity of comment sections suggest that some users have a need to comment and is gratified by it. And so, the removal of comment sections from a news site would make some users less satisfied with it.

As we saw in the literature review, a 2020 report stated that 25% of those who had previously commented on a news site, became more negative towards the news site when comment sections were closed, and that the average time spent on the site was reduced (Stroud, Murray, and Kim 2020). This means that the news sites, who depend on having readers to make a profit, have an incentive to keep comment sections, as long as hosting and moderating them doesn't become too much of a burden. But that is a financial incentive for news sites and says little about the role of comment sections in a democracy apart from being a potential financial support for a democratically valuable institution. However, there is a democratic argument to be made for why having open comment sections could be valuable.

Comment sections keep people engaged with the news media, and closing them down could create more distance between the news media and its audience (Williams and Sebastian 2022). If the media is an important establishment in a democratic society, then it would be desirable to have the citizens of that society engaging with the media to become better informed. And if hosting comment sections, toxic as they may be, is what it takes to keep some people coming back for more professional news, then comment sections would be serving a positive role. If comment sections are closed, those individuals who come to news sites to comment, could instead find an outlet for their opinions on alternative news sites and social media that are more plagued with echo chambers and filter bubbles. As the writer Sandra Newman writes:

> Comments sections host people of every political orientation, intelligence level, and psychiatric diagnosis. You encounter every talking point you hate, expressed with gloating certainty and an ear-shattering disregard for grammar. Some of the commenters could fairly be described as idiots. […] But it's not an echo chamber. On Twitter or Facebook, when you discuss climate change or same-sex marriage, you're talking to a self-selecting audience that mostly already agrees with you. In the comments section, you're talking to anyone who has Internet access. (Newman 2015)

Newman makes an interesting point about comment sections. Even though theories about echo chambers and filter bubbles have come under criticism (Bruns 2019, 3), these concepts have been considered a substantial challenge in the modern, digital age (Pariser 2011; Flaxman, Goel, and Rao 2016). Whether the societal threat of filter bubbles and echo chambers is substantial or not, the comment section of a mainstream news site is not where you would find such bubbles and chambers. And so, comment sections may serve a role as a meeting place, where the varied audience of a news site, with their differing political believes and opinions, are at least challenged and exposed to differing opinions.

6.6. Limitations

6.6.1 Sampling

As with any research project, this thesis has limitations that are worth discussing. Several of these were discussed in the chapter on methodology but are worth repeating here as I consider the challenges met by any researcher studying comment sections.

Firstly, the sampling process can be challenging. The researcher must ensure they gather all relevant comments. If the sampling process takes place too early, the researcher might miss later additions to the debate. In my case, I sampled comments written months earlier. This meant that enough time had passed that comments may have been deleted for one reason or another. It is unclear how any deleted comments may have influenced the results of the studies presented in this thesis, but it is nevertheless something to be aware of.

6.6.2. Data from a specific time and place

Continuing the discussion on how sampling have affected the results presented in this thesis, I must acknowledge the limitations when collecting data from a specific place and time. The comments analyzed for this thesis were all collected from American news sites and were written in 2018 and 2019. This has serious implications on the generalization of the results, in that they may not be generalizable at all. The political climate in the U.S. in 2018-19 was not the same as it was just a few years earlier, and certainly not the same as in many other countries around the world. All countries have specific issues occupying the interests of its population, and these issues change over time. As such, it is difficult to generalize results from one country at one specific point in time. Therefore, it would be accurate to say that this thesis is not about comment sections in general, but a case study of comments on American mainstream media in the late 2010’s.

The fact that this thesis has used comments from mainstream media is worth emphasizing. With the spread of the internet, alternative media has become popular. While it would be interesting to compare mainstream and alternative media in all the research projects presented in this thesis, it was outside the scope of what this thesis set out to do. In any case, the fact that the comments used for my research projects are only from mainstream media makes generalization of the results more difficult.

6.6.3. Coding scheme

In relation to my article, *Anonymity and Inhibition in Newspaper Comments (M. Knustad and Johansson 2021),* it is worth noting that the exact number of uncivil comments found in any research project is dependent upon not only the data, but how that data is interpreted. For my own study I used a coding scheme developed by Papacharissi (2004). While this was sufficient to look for differences between anonymous and non-anonymous comments, as any faults in the coding scheme would apply to both groups, one should be careful not to make definitive statements about the total number of uncivil comments based on my research – only that uncivil comments are prevalent. If the coding scheme is too liberal in what it defines as uncivil or impolite, the number of uncivil or impolite comments could be inflated. However, as we've seen in the literature review, my research is not the only one to find a great number of uncivil or impolite comments, and it is certainly a problem.

6.6.4. What are we actually researching?

The final limitation of this study is in my mind the most important one. When researching comments, researchers tend to frame their research a certain way, and I am certainly guilty of this as well. Researchers write about how people behave in comment sections and design research methodologies to uncover an aspect of online human behavior. Often this is done with quantitative research designs, where a vast number of comments are analyzed. There is a problem with this line of thinking, however. Researchers often study published comments, long after they were written. But researchers, including myself, rarely have access to comments that were not published or deleted after publication by moderators. In other words, the data being studied is not all comments *written*, but all comments *published*.

By assuming that all published comments are equal to all written comments, researchers are victim of survivorship bias (Clements and Bullivant 2022), whereby they're unaware that the analyzed comments are those that have survived the filtering process that is moderation. Of course, I can't assume that all researchers studying comment sections are unaware of this problem. But speaking for myself, it is an issue that is easy to forget when one is investigating how people behave online.

The survivorship bias becomes especially problematic when studying differences between platforms, since different platforms may have different moderation policies. This is something that I have previously discussed when writing about observed differences between Politico, The Washington Post and The New York Times. Therefore, any such differences should be taken with a grain of salt. Furthermore, the survivorship bias is especially relevant when studying toxicity, as toxic comments are precisely the comments that are likely to be deleted.

In the end, researchers must acknowledge that when they sample data from platforms such as news sites, they are not sampling all produced data. They are sampling all published data. Therefore, it is important to remember that when studying comment sections, we are not studying written comments, but published comments. We are not necessarily studying online human behavior, but human behavior that has been found acceptable by moderators.

7. Conclusion

The role of comment sections in a democratic society is unclear, but certainly challenging. The greatest obstacle for comment sections playing an important, positive role is the prevalence of toxic disinhibition. A substantial number of comments are toxic, as demonstrated by numerous studies, including my own. It is important to note, however, that the majority of comments are not toxic, meaning that most commenters are at the very least civil in their discussions. Of course, a comment not being toxic does not mean that it is a valuable contribution to a debate – just that it is not a toxic one. Comment sections, being such public platforms, are also susceptible to trolling. Even as the public become aware of trolling, my research shows that comment sections do not self-regulate by ignoring or combating the influence of perceived trolls, and that accusations of trolling are used to delegitimize opposing political arguments.

Given the lack of deliberation, susceptibility to trolling and toxicity found in comment sections, it is difficult to argue that they have a positive role in society when judged by the standards of most democratic theories. If, however, one considers comment sections as an anti-establishment or non-professional platform, then the role of comment sections could be argued to be as an opposing force to the established, professional journalistic field. Therefore, the most valuable role of comment sections could be to serve as a forum for critique of the media. But for such critique to be useful for the media, meaning that it can be used by the media to improve the journalistic work, it must be better understood. The research in this thesis has resulted in a two-dimensional coding scheme that has unveiled more detailed information about media criticism and may be useful for future research and the development of systems to categorize critical comments. Comment sections may also have a role as an incentive for people who might not otherwise visit news sites to engage with mainstream media, which would expose them to different opinions and professionally published news.

Finally, I find that comment sections have a useful role as a topic of research and discussion into the limits and expectations of freedom of speech and public discourse. Comment sections provide us with a unique forum for considering what freedom of speech is in practice, how it relates to the treatment of others, and where we draw the line between what is acceptable to say or not. They force us to think about what one should have to endure when expressing oneself publicly, and what it is reasonable to expect someone to endure. Therefore, in the end I find that the most practical and useful role comment sections have in a democratic society, is as a topic of debate and research. Not just debating and researching the comment sections themselves, but how free speech is and should be practiced. While I cannot make any definitive conclusions about this broad and important topic, I encourage further discussion about not just the role of comment sections in a democratic society, but the role of speech itself.

Comment sections are not going anywhere anytime soon. Despite several news sites having closed their comment sections, most seem to have no plans to remove them. While comment sections are evolving and newspapers are periodically updating their technologies with the hope of providing a better forum for reader engagement, it seems that the causes of toxic disinhibition are difficult to both identify and combat. It is not for me to judge if the fight against toxic comments is worth it, or if newspapers should consider new alternatives for facilitating discussion. I do find that comment sections have great potential if done right. It is just very difficult to determine how one might create a sufficiently civil and valuable comment section, and it is not unreasonable to think that other solutions should be considered. Either way, internet technologies will continue to provide people with different channels for public expression, both good and bad. If comment sections will continue to be one of those channels in the future, remains to be seen.

8. Bibliography

Anderson, Ashley A., Dominique Brossard, Dietram A. Scheufele, Michael A. Xanos, and Peter Ladwig. 2014. "The “nasty effect:” Online incivility and risk perceptions of emerging technologies." *Journal of Computer-Mediated Communication* 19 (3): 373-383.

Anderson, Ashley A., Sara K. Yeo, Dominique Brossard, Dietram A. Scheufele, and Michael A. Xanos. 2018. "Toxic Talk: How Online Incivility Can Undermine Perceptions of Media." *International Journal of Public Opinion Research* 30 (1). https://doi.org/10.1093/ijpor/edw022.

Anderson, C.W. 2012. "From Indymedia to Demand Media." In *Social Media Reader*, edited by Michael Mandiberg. New york: New York University Press.

Artime, Michael. 2016. "Angry and Alone: Demographic Characteristics of Those Who Post to Online Comment Sections." *Social Sciences* 5 (4). https://doi.org/10.3390/socsci5040068.

Bae, Mikyeung. 2016. "The effects of anonymity on computer-mediated communication: The case of independent versus interdependent self-construal influence." *Computers in Human Behavior* 55: 300-309. https://doi.org/10.1016/j.chb.2015.09.026.

Barlett, Christopher P., Douglas A. Gentile, and Chelsea Chew. 2016. "Predicting cyberbullying from anonymity." *Psychology of Popular Media Culture* 5 (2): 171-180. https://doi.org/10.1037/ppm0000055.

Bastos, Marco T., and Dan Mercea. 2019. "The Brexit Botnet and User-Generated Hyperpartisan News." *Social Science Computer Review* 37 (1): 38-54. https://doi.org/10.1177/0894439317734157.

Benson, Lee, Ira Richard Harkavy, and John L. Puckett. 2007. *Dewey's dream : universities and democracies in an age of education reform : civil society, public schools, and democratic citizenship*. Philadelphia: Temple University Press.

Berg, Janne. 2016. "The impact of anonymity and issue controversiality on the quality of online discussion." *Journal of Information Technology & Politics* 13 (1): 37-51. https://doi.org/10.1080/19331681.2015.1131654.

Bilton, Ricardo. 2014. "Why some publishers are killing their comment sections." *Digiday* (blog). https://digiday.com/media/comments-sections/.

Bishop, J. D. 1970. "The Cleroterium." *J. Hell. Stud* 90: 1-14. https://doi.org/10.2307/629749.

Blom, Robin, Serena Carpenter, Brian J. Bowe, and Ryan Lange. 2014. "Frequent Contributors Within U.S. Newspaper Comment Forums." *American Behavioral Scientist* 58 (10): 1314-1328. https://doi.org/10.1177/0002764214527094.

Boehm, Ryan. 2012. Agora. Hoboken, NJ, USA: Hoboken, NJ, USA: John Wiley & Sons, Inc.

Broniatowski, David A., Amelia M. Jamison, Sihua Qi, Lulwah Alkulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. 2018. "Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate." *American journal of public health* 108 (10): 1378-1384. https://doi.org/10.2105/AJPH.2018.304567.

Bruns, Axel. 2005. *Gatewatching: collaborative online news production*. New York: Peter Lang Publishing, Inc.

---. 2019. *Are Filter Bubbles Real?* Cambridge: Polity Press.

Buckels, Erin E., Paul D. Trapnell, and Delroy L. Paulhus. 2014. "Trolls just want to have fun." *Personality and Individual Differences* 67: 97-102.

Casselberry, Samuel E. 1971. "ALBERT MEHRABIAN." *American Anthropologist* 75: 1926-1927.

Cheng, S. L., W. H. Lin, F. K. Phoa, J. S. Hwang, and W. C. Liu. 2015. "Analysing the Unequal Effects of Positive and Negative Information on the Behavior of Users of a Taiwanese On-Line Bulletin Board." *PLoS One* 10 (9). https://doi.org/10.1371/journal.pone.0137842.

Chiou, W. B. 2006. "Adolescents' sexual self-disclosure on the internet: Deindividuation and impression management." *Adolescence* 41 (163).

Clements, Ben, and Stephen Bullivant. 2022. "Why Younger Catholics Seem More Committed: Survivorship Bias and/or “Creative Minority” Effects among British Catholics." *Journal for the scientific study of religion* 61 (2): 450-475. https://doi.org/10.1111/jssr.12791.

Coe, Kevin, Kate Kenski, and Stephen A. Rains. 2014. "Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments." *Journal of Communication* 64 (4): 658-679. https://doi.org/10.1111/jcom.12104.

Crick, Bernard. 2002. *Democracy: A very short introduction*. Kindle ed. New York: Oxford University Press.

Curran, James. 2011. *Media and Democracy*. New York: Routledge.

de Seta, Gabriele. 2018. "Trolling, and Other Problematic Social Media Practices." In *The SAGE Handbook of Social Media*, edited by Jean Burgess, Alice Marwick and Thomas Poell. UK: SAGE Publications Ltd.

Dewey, John. 2003. "The Eclipse of the Public." In *Civil Society Reader*, edited by Virginia Hodgkinson and Michael W. Foley. London: Tufts University Press.

Dimock, Michael. 2019. "An update on our research into trust, facts and democracy." Pew Research Center. Last Modified June 5. Accessed June 17. https://www.pewresearch.org/2019/06/05/an-update-on-our-research-into-trust-facts-and-democracy/.

Domingo, David, Thorsten Quandt, Ari Heinonen, Steve Paulussen, Jane B. Singer, and Marina Vujnovic. 2008. "PARTICIPATORY JOURNALISM PRACTICES IN THE MEDIA AND BEYOND: An international comparative study of initiatives in online newspapers." *Journalism Practice* 2 (3): 326-342. https://doi.org/10.1080/17512780802281065.

Edgarly, Stephanie, Emily Vraga, Timothy Fung, Tae Joon Moon, Woo Hyun Yoo, and Aaron Veenstra. 2009. "YouTube as a public sphere: The Proposition 8 debate." The Association of Internet Researchers Conference, 2009-10-8.

Elgesem, Dag, and Tomas Vie Nordeide. 2016. "Anonymity and tendentiousness in online newspaper debates." In *Journalism Re-examined. Digital Challenges and Professional Reorientations. Lessons from Northern Europe*, edited by Martin Eide, Helle Sjøvaag and Leif Ove Larsen. Bristol: Intellect Books.

Ellis, Justin. 2015. "What happened after 7 news sites got rid of reader comments." *Nieman Lab* (blog), *Neiman Lab*. https://www.niemanlab.org/2015/09/what-happened-after-7-news-sites-got-rid-of-reader-comments/.

Faden, Ruth R., and Tom L. Beauchamp. 1986. *A History and Theory of Informed Consent*. New York: Oxford University Press.

Ferree, Myra Marx, William A. Gamson, Jürgen Gerhards, and Dieter Rucht. 2002. "Four Models of the Public Sphere in Modern Democracies." *Theory and Society* 31: 289-324.

Finley, Klint. 2015. "A brief history of the end of the comments." *Wired* (blog). https://www.wired.com/2015/10/brief-history-of-the-demise-of-the-comments-timeline/.

Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. "Filter bubbles, echo chambers, and online news consumption." *Public Opinion Quarterly* 80 (Special Issue): 298-320.

Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora. 2021. De nasjonale forskningsetiske komiteene.

franzke, aline shakti, Anja Bechmann, Michael Zimmer, Charles M. Ess, and the Association of Internet Researchers. 2020. *Internet research: Ethical Guidelines 3.0.* (Association of Internet Researchers). https://aoir.org/reports/ethics3.pdf.

Frischlich, Lena, Svenja Boberg, and Thorsten Quandt. 2019. "Comment Sections as Targets of Dark Participation? Journalists' Evaluation and Moderation of Deviant User Comments." *Journalism Studies* 20 (14): 2014-2033. https://doi.org/10.1080/1461670X.2018.1556320.

Geiger, R. Stuart. 2009. "Does Habermas Understand the Internet? The Algorithmic Construction of the Blogo/Public Sphere." *Gnovis, A Journal of Communication, Culture, and Technology* 10 (1): 1-29.

Gillespie, Tarleton. 2018. *Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media*. New Haven: Yale University Press.

Gilovich, Thomas, Sacher Keltner, Serena Chen, and Richard E. Nisbett. 2016. *Social Psychology*. London: W.W. Norton & Company Ltd.

Goffman, Erving. 1956. "The Presentation of Self in Everyday Life." Department of Social Anthropology, University of Edinburgh.

Gonçalves, João. 2015. "A peaceful pyramid? Hierarchy and anonymity in newspaper comment sections." *Observatorio* 9 (4): 1-13.

Graham, Brett. 2016. "We're totally getting married...: Verbal Irony use in Computer-mediated Communication." Bachelor thesis, Southern Illenois University.

Graham, Todd, and Scott Wright. 2015. "A Tale of Two Stories from "Below the Line"." *The International Journal of Press/Politics* 20 (3): 317-338. https://doi.org/10.1177/1940161215581926.

Gripsrud, Jostein. 2017. *Allmenningen: Historien om norsk offentlighet*. Oslo: Universitetsforlaget.

Gutmann, Amy, and Dennis Thompson. 2004. *Why Deliberative Democracy*. Princeton: Princeton University Press.

Habermas, Jürgen. 1991. *The Structural Transformation of the Public Sphere*. United States of America: MIT Press.

Hall, Stuart. 2006. "Encoding/Decoding." In *Media and cultural studies*, edited by Meenakshi Gigi Durham and Douglas M. Kellner, 163-173. Malden: Blackwell Publishing.

Halliday, Josh. 2011. "SXSW 2011: 4Chan founder Christopher Poole on anonymity and creativity." *The Guardian*, March 13, 2011. https://www.theguardian.com/technology/2011/mar/13/christopher-poole-4chan-sxsw-keynote-speech.

Hardaker, Claire. 2015. "'I refuse to respond to this obvious troll': an overview of responses to (perceived) trolling." *Corpora* 10 (2): 201-229. https://doi.org/10.3366/cor.2015.0074.

Hirsh, Jacob B., Adam D. Galinsky, and C. B. Zhong. 2011. "Drunk, Powerful, and in the Dark: How General Processes of Disinhibition Produce Both Prosocial and Antisocial Behavior." *Perspectives on Psychological Science* 6 (5): 415-427.

Hogan, Bernie. 2013. "Pseudonyms and the Rise of the Real-Name Web." In *A Companion to New Media Dynamics*, edited by John Hartley, Jean Burgess and Axel Bruns, 290-308. Chichester: Blackwell publishing Ltd.

Hubler, Mike T., and Diana Calhoun Bell. 2003. "Computer-mediated humor and ethos: Exploring threads of constitutive laughter in online communities." *Computers and composition* 20: 277-294.

Hume, Tim. 2013. "How FBI caught Ross Ulbricht, alleged creator of criminal marketplace Silk Road." *CNN*, October 5, 2013. https://edition.cnn.com/2013/10/04/world/americas/silk-road-ross-ulbricht/index.html.

Jacobsen, Catrine, Toke Reinholt Fosgaard, and David Pascual-Ezama. 2018. "Why Do We Lie? A Practical Guide to the Dishonesty Literature." *Journal of Economic Surveys* 32 (2): 357-387. https://doi.org/10.1111/joes.12204.

Jessup, Leonard M., Terry Connolly, and Jolene Galegher. 1990. "The Effects of Anonymity on DDSS Group Process with an idea-Generating Task." *MIS Quarterly* 14 (3).

Johnson, Steven. 2016. *Wonderland: How Play Made the Modern World*. London: Penguin Group.

Joinson, Adam N. 2001. "Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity." *European Journal of Social Psychology* 31: 177-192.

Kirkpatrick, David. 2011. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. New York: Simon & Schuster Paperbacks.

Knustad, Magnus. 2020. "Get lost, troll: How accusations of trolling in newspaper comment sections affect the debate." *First Monday* 25 (8). https://doi.org/https://doi.org/10.5210/fm.v25i8.10270.

Knustad, Magnus André. 2018. How platform affects comments on news articles. A qualitative analysis of comments from a newspaper's comment section and Facebook page. The University of Bergen.

Knustad, Magnus, and Christer Johansson. 2021. "Anonymity and Inhibition in Newspaper Comments." *Information (Basel)* 12 (3): 106. https://doi.org/10.3390/info12030106.

Korslien, Hansine. 2014. "Ytringsfrihet fra folkedypet." *VG*, January 30, 2014. https://www.vg.no/nyheter/meninger/i/MKBMK/ytringsfrihet-fra-folkedypet.

Kruger, Justin, Nicholas Epley, Jason Parker, and Zhi-Wen Ng. 2005. "Egocentrism over e-mail: Can we communicate as well as we think?" *Journal of Personality and Social Psychology* 89 (6): 925-936.

Laclau, Ernesto. 2005. *On Populist Reason*. New York and London: Verso.

Lapidot-Lefler, Noam, and Azy Barak. 2012. "Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition." *Computers in Human Behavior* 28 (2): 434-443. https://doi.org/10.1016/j.chb.2011.10.014.

Lebeuf, Carlene, Margaret-Anne Storey, and Alexey Zagalsky. 2018. "Software Bots." *IEEE Software* 35 (1): 18-23. https://doi.org/10.1109/MS.2017.4541027.

Li, Xigen. 2010. *Internet Newspapers: The Making of a Mainstream Medium*. New York: Routledge.

Løvlie, Anders Sundnes. 2018. "Constructive Comments?: Designing an online debate system for the Danish Broadcasting Corporation." *Journalism Practice* 12 (6): 781-798. https://doi.org/10.1080/17512786.2018.1473042.

Løvlie, Anders Sundnes, Karoline Andrea Ihlebæk, and Anders Olof Larsson. 2017. "User Experiences with Editorial Control in Online Newspaper Comment Fields." *Journalism Practice* 12 (3): 362-381. https://doi.org/10.1080/17512786.2017.1293490.

Lunt, Peter, and Sonia Livingstone. 2013. "Media studies' fascination with the concept of the public sphere: critical reflections and emerging debates." *Media, Culture and Society* 35 (1): 87-96.

Mansbridge, Jane, James Bohman, Simone Chambers, Thomas Christiano, Archon Fung, John Parkinson, Dennis F. Thompson, and Mark E. Warren. 2012. "A systemic approach to deliberative democracy." In *Deliberative Systems: Deliberative Democracy at the Large Scale*, edited by John Parkinson and Jane Mansbridge. New York: Cambridge University Press.

Markham, Annette, and Elizabeth Buchanan. 2012. "Ethical Decision Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)." Association of Internet Researchers. Accessed August 20. https://aoir.org/reports/ethics2.pdf.

Marwick, Alice E., and Danah Boyd. 2010. "I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience." *New Media & Society* 13 (1): 114-133. https://doi.org/10.1177/1461444810365313.

Mead, George Herbert. 1934. *Mind, Self, and Society*. Chicago: University of Chicago Press.

Medietilsynet. August 2019. *Kritisk medieforståelse i den norske befolkningen - En undersøkelse fra Medietilsynet - Delrapport 3: Debattdeltakelse i media.* Medietilsynet. https://www.medietilsynet.no/globalassets/publikasjoner/2020/kritisk-medieforstaelse-samlerapport-og-delrapporter/delrapport-3-kmf-debattdeltakelse-i-media.pdf.

Mehrabian, Albert. 1971. *Silent Messages: Implicit Communication of Emotions and Attitudes*. Belmont: Wadsworth Publishing Company.

Merriam-Webster. 2019a. comment. In *Merriam-Webster*.

---. 2019b. forum. In *Merriam-Webster*.

---. 2019c. pub. In *Merriam-Webster*.

---. 2023. role. In *Merriam Webster*.

Misoch, Sabina. 2015. "Stranger on the internet: Online self-disclosure and the role of visual anonymity." *Computers in Human Behavior* 48: 535-541. https://doi.org/10.1016/j.chb.2015.02.027.

Mouffe, Chantal. 2018. *For a Left Populism*. New York and London: Verso.

---. 2022. *Towards a Green Democratic Revolution: Left populism and the Power of Affects*. New York and London: Verso.

Newman, Sandra. 2015. "In Defense of Comment Sections." *Slate*, November 4.

Papacharissi, Zizi. 2004. "Democracy online: civility, politeness, and the democratic potential of online political discussion groups." *new media & society* 6 (2): 259-283. https://doi.org/10.1177/1461444804041444.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet is Hiding from You*. New York: Penguin Press.

Perote-Peña, Juan, and Ashley Piggins. 2015. "A Model of Deliberative and Aggregative Democracy." *Economics and Philosophy* 31 (1): 93-121. https://doi.org/10.1017/S0266267114000418.

Philips, D.J. 1996. "Defending the boundaries: Identifying and countering threats in a Usenet newsgroup." *The Informational Society* 12 (1): 39-62.

Pieper, Ian, and Colin Thomson. 2016. "Beneficence as a principle in human research." *Monash Bioethics Review* 34 (2): 117-135. https://doi.org/10.1007/s40592-016-0061-3.

Porta, Miquel. 2014. Observational epidemiological study. In *A Dictionary of Epidemiology*. New York: Oxford University Press.

Postmes, Tom, Russell Spears, Khaled Sakhel, and Daphne de Groot. 2001. "Social Influence in Computer-Mediated Communication: The Effects of Anonymity on Group Behavior." *Personality and Social Psychology Bulletin* 27 (10): 1243-1254.

Quandt, Thorsten. 2018. "Dark Participation." *Media and Communication* 6 (4): 36. https://doi.org/10.17645/mac.v6i4.1519.

Ramnefjell, Geir. 2016. "Dagbladets kommentarfelt (1996 - 2016)." Dagbladet.no. Last Modified 2016-01-28.

Reagle, Joseph M. 2015. *Reading the comments: Likers, Haters, and Manipulators at the Bottom of the Web*. Sabon: MIT Press.

Rettberg, Jill Walker. 2014. *Blogging*. 2 ed. Cambridge: Polity Press.

Rhodes, P. J. 2012. Kleroterion. Hoboken, NJ, USA: Hoboken, NJ, USA: John Wiley & Sons, Inc.

Riffe, Daniel, Stephen Lacy, and Frederick Fico. 2014. *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. New York: Routledge.

Rösner, Leonie, and Nicole C. Krämer. 2016. "Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments." *Social Media + Society*: 1-13. https://doi.org/10.1177/2056305116664220.

Rossini, Patrícia. 2019. "Toxic for Whom? Examining the Targets of Uncivil and Intolerant Discourse in Online Political Talk." In *Voices: Exploring the shifting contours of communication*, edited by Patricia May and Donald Matheson. New York: Peter Lang Publishing Inc.

Rowe, Ian. 2014. "Civility 2.0: a comparative analysis of incivility in online political discussion." *Information, Communication & Society* 18 (2): 121-138. https://doi.org/10.1080/1369118x.2014.940365.

Santana, Arthur D. 2014. "Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards." *Journalism Practice* 8 (1): 18-33.

---. 2019. "Toward quality discourse: Measuring the effect of user identity in commenting forums." *Newspaper Research Journal* 40 (4): 467-486. https://doi.org/10.1177/0739532919873089.

Savill, Richard. 2000. "Harry Potter and the mystery of J K's lost initial." *The Telegraph*, July 19, 2000. https://www.telegraph.co.uk/news/uknews/1349288/Harry-Potter-and-the-mystery-of-J-Ks-lost-initial.html.

Scott, Craig R. 2012. "Benefits and Drawbacks of Anonymous Online Communication: Legal Challenges and Communicative Recommendations." *Free Speech Yearbook* 41 (1): 127-141. https://doi.org/10.1080/08997225.2004.10556309.

Sennett, Richard. 1986. *The Fall of Public Man*. London: Faber and Faber.

Shachaf, Pnina, and Noriko Hara. 2010. "Beyond vandalism: Wikipedia trolls." *Journal of Information Science* 36 (3): 357-370. https://doi.org/10.1177_0165551510365390.

Shafe, James. 2014. "Challenges for a revised view of Bentham on public reasoning." *Revue d'études benthamiennes* (13). https://doi.org/10.4000/etudes-benthamiennes.761.

Singer, Jane B., Steve Paulussen, and Alfred Hermida. 2011. *Participatory journalism : guarding open gates at online newspapers*. Malden, Mass: Wiley-Blackwell.

Smith, John. 1965. "Message, Meaning and Context in Ethology." *The American Naturalist* 99 (908): 405-409.

Solheim, John, and Trine Syvertsen. 2021. "Norsk presses historie." Last Modified November 19. 2021. Accessed September 12. https://snl.no/norsk_presses_historie.

Stalsberg, Linn. 2015. "Det problematiske kommentarfeltet." *VG*, June 25, 2015. https://www.vg.no/nyheter/meninger/i/mkQrq/det-problematiske-kommentarfeltet.

Stein, Edward. 2003. "Queers anonymous: Lesbians, gay men, free speech, and cyberspace." *Harvard Civil Rights - Civil Liberties Law Review* 38 (1): 159-213.

Strömbäck, Jesper. 2005. "In Search of a Standard: four models of democracy and their normative implications for journalism." *Journalism Studies* 6 (3): 331-345. https://doi.org/10.1080/14616700500131950.

Stroud, Natalie Jomini, Ashley Muddiman, and Joshua M. Scacco. 2016. "Like, recommend, or respect? Altering political behavior in news comment sections." *new media & society*: 1-17. https://doi.org/10.1177/1461444816642420.

Stroud, Natalie Jomini, Caroline Murray, and Yujin Kim. 2020. *News comments: What happens when they're gone or when newsrooms switch platforms.* Center for Media Engagement. https://mediaengagement.org/research/comment-changes/.

Stroud, Natalie Jomini, Emily Van Duyn, and Cynthia Peacock. 2016. *Survey of Commenters and Comment Readers.* Center for Media Engagement. https://mediaengagement.org/research/survey-of-commenters-and-comment-readers/.

"Studer digital kultur." 2016. Universitetet i Bergen. Accessed April 18. https://www.uib.no/fag/digitalkultur/96746/studer-digital-kultur.

Suler, John. 2005. "The Online Disinhibition Effect." *International Journal of Applied Psychoanalytic Studies* 2, no. 2: 184-188.

---. 2016. *Psychology of the Digital Age: Humans Become Electric*. New York: Cambridge University Press.

Suler, John, and W. L. Philips. 1998. "The Bad Boys of Cyberspace: Deviant Behavior in a Multimedia Chat Community." *Cyberpsychology & Behavior* 1 (3): 275-294.

Toepfl, Florian, and Eunike Piwoni. 2015. "Public Spheres in Interaction: Comment Sections of News Websites as Counterpublic Spaces." *Journal of Communication* 65 (3): 465-488. https://doi.org/10.1111/jcom.12156.

Vergeer, Maurice. 2015. "Twitter and Political Campaigning." *Sociology Compass* 9 (9): 745-760. https://doi.org/10.1111/soc4.12294.

Wallsten, Kevin, and Melinda Tarsi. 2015. "Persuasion from Below?" *Journalism Practice* 10 (8): 1019-1040. https://doi.org/10.1080/17512786.2015.1102607.

Wang, Hongjie, and Hong Yan. 1996. "Flaming: More Than a Necessary Evil for Academic Mailing Lists." *Electronic Journal of Communication* 6 (1).

Whiting, Anita, and David Williams. 2013. "Why people use social media: a uses and gratifications approach." *Qualitative market research* 16 (4): 362-369. https://doi.org/10.1108/QMR-06-2013-0041.

Williams, Kat, and Bailey Sebastian. 2022. "Online Comment Sections: Does Taking Them Down Enhance or Hurt Dialogue in a Democracy?" *Journal of Media Ethics* 37 (4): 285-287. https://doi.org/doi.org/10.1080/23736992.2021.1976645.

Wingenbach, Edward C. 2011. *Institutionalizing agonistic democracy : post-foundationalism and political liberalism*. Burlington, Vt.: Ashgate.

Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109 (1): 121-136.

Zimmerman, Adam G., and Gabriel J. Ybarra. 2016. "Online aggression: The influences of anonymity and social modeling." *Psychology of Popular Media Culture* 5 (2): 181-193. https://doi.org/10.1037/ppm0000038.

9. Articles

Knustad, M., Johansson, C. (2021) Anonymity and Inhibition in Newspaper Comments. *Information, 12(3):106.* https://doi.org/10.3390/info12030106

Knustad, M. (2020). Get Lost, Troll: How Accusations of Trolling in Newspaper Comment Sections Effect the debate. *First Monday, 25(8).* https://doi.org/10.5210/fm.v25i8.10270

Knustad, M. Media criticism in newspaper comment sections: Do comment sections constitute a democratically valuable forum for constructive criticism of the media?

information

Article

Anonymity and Inhibition in Newspaper Comments

Magnus Knustad * and Christer Johansson *

Department of Linguistic, Literary and Aesthetic Studies, University of Bergen, 5007 Bergen, Norway
* Correspondence: Magnus.Knustad@uib.no (M.K.); Christer.Johansson@uib.no (C.J.)

Abstract: Newspaper comment sections allow readers to voice their opinion on a wide range of topics, provide feedback for journalists and editors and may enable public debate. Comment sections have been criticized as a medium for toxic comments. Such behavior in comment sections has been attributed to the effect of anonymity. Several studies have found a relationship between anonymity and toxic comments, based on laboratory conditions or the comparison of comments from different sites or platforms. The current study uses real-world data sampled from *The Washington Post* and *The New York Times*, where anonymous and non-anonymous users comment on the same articles. This sampling strategy decreases the possibility of interfering variables, ensuring that any observed differences between the two groups can be explained by anonymity. A small but significant relationship between anonymity and toxic comments was found, though the effects of both the newspaper and the direction of the comment were stronger. While it is true that non-anonymous commenters write fewer toxic comments, we observed that many of the toxic comments were directed at others than the article or author of the original article. This may indicate a way to restrict toxic comments, while allowing anonymity, by restricting the reference to others, e.g., by enforcing writers to focus on the topic.

Keywords: anonymity; inhibition; disinhibition; incivility; toxic comments

Citation: Knustad, M.; Johansson, C. Anonymity and Inhibition in Newspaper Comments. *Information* **2021**, *12*, 106. https://doi.org/10.3390/info12030106

Academic Editor: Diego Reforgiato Recupero

Received: 19 January 2021
Accepted: 27 February 2021
Published: 3 March 2021

1. Introduction

Comment sections are a common feature of online news sites and have been described as a staple of the online experience [1]. Among their functions are to engage readers, to provide a democratic voice to the audience, to document the popularity of the source, and to provide journalists and editors with direct feedback on their published articles. Such information can obviously be very valuable, and based on the number of news sites that host comment sections, news sources want to have this interaction with their audience.

In the past years there has been a movement against anonymous online content, based on the assumption that anonymity leads to hostility and insults, and allows for cyberbullying [2]. To combat unwanted comments, many news sites have adopted a policy that requires commenters to use their real names when commenting. Many news sites do this by using a Facebook plugin, where users must log in to their Facebook account to comment on news articles. Some sites have gone even further by closing their comment sections and use their Facebook pages as the primary platform for user engagement and commenting [3–6], raising concerns about privacy [7].

Many academics studying comment sections are, understandably, focused on the negative aspects of commenting, specifically toxic disinhibition in comment sections. Toxic disinhibition is defined by Suler [8] as online behavior that is rude, critical, angry, hateful and threatening. Based on this definition, this article will use the term *toxic comments* when referring to the subject matter of this study. This included the literature review, where referenced studies may have used other terms. Other researchers have used terms such as *uncivil* and *impolite* comments [9]. These two terms are used in the coding scheme developed by Papacharissi [10] that was used in the current study.

Several studies have investigated why seemingly normal people behave in a disinhibited way when commenting, and anonymity is often brought up as an explanatory factor for toxic disinhibition in comment sections [9–12]. There is, however, a methodological problem when studying comment sections. To compare anonymous and non-anonymous communication, some researchers have used laboratory settings [12,13]. While experimental research designs can provide important insights, there is always the question of generalizing results to real-world situations. Other researchers have studied the differences between anonymous and non-anonymous comment sections on different news sites [11], which means that the two experimental groups come from different populations. With such sampling strategies one might risk results being affected by other variables than anonymity. Some researchers have even tried to study the effect of anonymity by comparing data from comment sections and other platforms, such as Facebook [9]. Obviously, there are many other variables than anonymity that could affect results when comparing a comment section to a social media platform.

The current study aims to improve the methodology of studying comments sampled from real-world sources. To study the effect of anonymity in real-world comment sections it is best to sample from platforms with both anonymous and non-anonymous commenters, and, thereby, estimate the effect of the platforms. This would amount to repeated measures on the same platforms, such that most other factors would be constant between platforms. Ideally the difference between the two groups would be whether they are anonymous or not. However, we must also account for individual differences since the same individuals may not comment both anonymously and openly, at least not under the same signature. The current study samples data from The Washington Post and The New York Times, two newspapers with a comment section where users can choose to use their real names or pseudonyms. The comments sampled from these newspapers represent anonymous and non-anonymous commenters on the same platforms. The research question for this study is: are anonymous comments more toxic than non-anonymous comments? In the literature review we see that the existing evidence for anonymity is based on experimental studies and studies where data is gathered from different platforms. In addition, there are other explanatory factors for why individuals may exhibit toxic disinhibition in comment sections. The null-hypothesis of this study is that there will not be a significant relationship between anonymity and toxicity. If there is an effect, how large is that effect?

There have been online communities as early as the ARPAnet, a precursor to the internet from 1969 [14]. In 1973, the Community Memory public bulletin board system was set up in Berkeley, and Internet was then viewed as a way to revitalize democracy and stimulate public debate and social change [15]. The World Wide Web in 1991, and the release of the Netscape Navigator in 1994, led to online editions of newspapers. By the year 2001 there were over 3400 online newspapers only in the U.S. [16]. At the same time, paper editions have declined.

Comment sections emerge as one form of participatory or constructive journalism [17]. Newspaper editors view comments as one of the most successful forms of audience interaction [18], and the intention is to continue supporting comment sections on online publications [19]. According to the Pew Research Center [20] about one in four Americans have contributed to comment sections. As many as 84% of newsreaders read comments, and studies have shown that reading comments can significantly affect readers' perception of public opinion, as well as change their personal opinion [21]. A more recent study found that news readers' perceptions of a news story was influenced more by the story itself than by comments made by other readers [22]. The same study found that the civility of comments did not influence readers' perception of the comment, but it did influence perceptions of the commenter and trust in the information. These findings suggest that how readers perceive and react to comments is a complex issue, but that there is an effect.

Comment sections have been criticized for being places of uncivil and impolite behavior. Papacharissi [10] developed a coding scheme for uncivil and impolite behavior in online forums, which Rowe [9] used to investigate the effect of anonymity in comment sections. Reported number of toxic comments vary from 4 to 22% [23]. The variation can be due to differences in which sections were studied, the definitions of toxic comments, and methodological differences. Different policies have also been shown to affect the number of uncivil comments [24].

Anonymity is defined by Scott [25] as "the condition in which a message source is absent or largely unknown to a message recipient". There are many reasons why someone would want to remain anonymous, according to Hogan [26]. External pressures may cause someone to express themselves anonymously in order to be treated as any individual, such as when female Victorian writers used male pseudonyms out of fear of being dismissed based on their gender. In a more modern example, the fantasy author Joanne Rowling published the Harry Potter books under the name J.K. Rowling because boys tend not to read books written by female authors [27]. She later published under the male pseudonym Robert Galbraith for a presumably different audience. Another reason for anonymity is internal motivations, where an individual has a desire to adopt a different persona. Functional motivations for anonymity are present when practical concerns dictate that a pseudonym is necessary, such as when other people share your name. Situational motivations arise when someone wants to keep different part of their online separate. Finally, there are personal motivations for anonymity, such as the desire to create an escape from everyday life [26], or when acting as a whistle-blower (cf. for example Wikileaks).

The study of anonymity and its effect on behavior has a long tradition in psychology. As early as 1895, Gustave LeBon studied how individuals take on a collective mindset when they are a part of a crowd, which makes them act differently than they do as individuals [28–30]. While not directly related to anonymity, LeBon's pioneering research was an impactful turning point in the history of psychological research, as it laid the groundworks for how socio-psychological processes could explain (unwanted) human behavior.

Over a half century later, researchers performed experiments to investigate the effects of anonymity. The term deindividuation was created to describe a state in which individuals experience a loss of their individual identity due to the anonymity provided by being in a large group [30]. Festinger, Pepitone and Newcomb [31] found that deindividuation caused by not feeling observed by others allowed test subjects to indulge in behavior from which they were usually restrained. The deindividualized test subjects made more negative comments about their parents than the control group, suggesting a relationship between the degree to which someone is identifiable and their willingness to make negative statements. To investigate deindividuation in real-world conditions, Diener, et al. [32] performed a study on Halloween where they observed if trick-or-treaters would steal candy or money when given the opportunity. Children who were not asked about who they were or where they lived, meaning that they remained anonymous, were more likely to steal. It was also found that children in groups were more likely to steal, pointing to both anonymity and crowd mentality as explanations for unwanted behavior. Modern theories of deindividuation, however, show that anonymity does not necessarily lead to antisocial behavior. Instead, anonymity has been found to lead to increased conformity with group norms, which again can lead to antisocial behavior depending on the norms of the group in any given situation [33]. The Social Identity/Deindividuation (SIDE) Model challenges traditional models of deindividuation that focus on the self being the basis of rational action and the group serving to impede the operation of such selfhood [34]. The SIDE-model emphasizes the effect of social identities on deindividuation. Reicher, Spears and Postmes [34] argues that anonymity within a group does not lead to uncontrolled behavior, but instead gives the members of a group the opportunity to "give full voice to their collective identities." This may also be negative, in that a group may more forcefully side against its opponents and inflate the sense of consensus on, and legitimacy of, the opinions within the group.

With computer-mediated communication came new opportunities for anonymity. It could be argued that online anonymity is an important requirement for our online lives. Having multiple identities when communicating with different people, such as friends, family or coworkers, is part of the human social experience. Throughout history it has been possible to share different identities depending on the social context [35]. According to role theory in social psychology we juggle different social roles, implying that having multiple personalities is a normal part of human nature. This is presumably true online as well [36]. As Hogan [26] points out when describing situational motivations for anonymity, when people use their real names it makes it possible for two completely different posts from different sources to be presented in the same search results. A pseudonym makes it possible to avoid context collapse, a phenomenon described by Marwick and Boyd [37] as an online situation where multiple audience flatten into one, making it impossible to differentiate self-representation strategies. In addition, contributors to forums at online news sites—especially the most frequent contributors—support anonymity, expressing positive views about how anonymity promotes freer and livelier conversation [38].

Anonymity involves not being held accountable for one's actions, which seems to underlie most concerns about anonymity [25,39,40]. Alongside alcohol consumption and social power, anonymity has a disinhibited effect that emerges from a common psychological mechanism; lower activation of the Behavioral Inhibition System [41]. These three factors may combine to escalate the effect of disinhibition. There is also a social factor to anonymity, in that other people being anonymous may lead to a person behaving in a toxic way [36]. Postmes et al. [42] found that anonymous group members are more likely to be affected by social influence, meaning that anonymous internet users are more likely to behave in an uncivil manner if others are uncivil.

Several studies have concluded that there is a relationship between toxic disinhibition and anonymity. Rowe [9] found that there was more incivility in comments on the Washington Post comment section than on the same articles on Facebook. This finding was explained by the fact that users of the Washington Post comment section are anonymous. This explanation, however, disregards other possible differences between the two platforms. While the Washington Post provides a standardized comment section, which allows for little functionality beyond commenting on articles, Facebook is a diverse social media platform where commenters do not even have to access an article to comment on it. In addition, it is not guaranteed that commenters on The Washington Post are anonymous. While the Washington Post allows for anonymous commenters, a commenter may also use his or her real name. Furthermore, when you make a comment on Facebook, all of the people in your friend list may not only potentially see your comment but be algorithmically directed towards it. This may restrict free expression, as writers know that someone who knows them may judge them. Obviously, even non-anonymous comments will tend to be more vapid on such a medium, possibly even ameliorated by a tendency to virtue signaling towards people you know socially. While Rowe's findings are interesting as a study of platforms, it is difficult to make definite conclusions about the effect of anonymity in general from his results. In another study, Rowe explores the deliberative value of comments on Facebook and The Washington Post, and concludes that comments left by website users were more deliberative than those left by Facebook users [43].

Dillon, Neo and Seely [44] found that comments from two news sites using a Facebook plugin were less civil and polite than those found on two news sites where commenters could comment anonymously. While this is an interesting result, it is possible that the results could be affected by the fact that the anonymous and non-anonymous comments were sampled from different sources. As the researchers point out in their discussion, "We did not take socio-democratic factors such as the political climate of geographical regions into consideration when choosing the four newspapers."

Santana [11] found a significant relationship between anonymity and civility when studying comments from three news sites allowing for anonymity and eleven news sites where commenters had to use their real names. Though, it is worth noting that Santana's results are based on studying anonymous and non-anonymous commenters in different populations. In another study, where 4800 comments were sampled from 30 news sites, Santana found that anonymous commenters were more likely to write uncivil comments [45].

While the higher number of sources compared to the three sites used in the 2014 study, the anonymous and non-anonymous comments are still sampled from different sites, and other interfering variables cannot be excluded.

The Huffington Post provides an interesting case study of anonymity. In its early days, the news site allowed users to comment using any chosen name. In December of 2013 the site changed its policy so that users had to authenticate their accounts through Facebook, while still allowing them to use a pseudonym to comment. In June of 2014, the site changed its policy again, this time implementing a Facebook plugin, meaning that users had to use their real names when commenting. In a large-scale study of comments from before and after the first change of policy—before and after they implemented a requirement of identification through Facebook in 2013—Fredheim, Moore and Naughton [46] found that comment quality improved after users had to authenticate their accounts. However, a similar study on the comment sections of Huffington Post [47] complicates the issue, as they found that the quality of commenting, measured by the cognitive complexity of comments, improved after the first change of policy, where users had to authenticate their accounts but could still use pseudonyms. However, the second change, when users had to use their real names when commenting, caused a decrease in the quality of discussions. Interestingly, after both reforms the quality of discussions improved over time. This indicates that the durability over time is a more important factor than whether using a real name or a pseudonym.

Lapidot-Lefler and Barak [12] found in an experimental research design that anonymity influenced the numbers of threats made by research participants. Anonymity was, however, not found to influence self-reported flaming, negative atmosphere or flaming-related expressions. Barlett, Gentile and Chew [48] used a longitudinal design involving questionnaires, and found that the more people feel that they are anonymous the more likely they are to cyberbully others. Zimmerman and Ybarra [13] found in an experimental research design that anonymous participants were more aggressive than those who were not anonymous. However, it is difficult to judge the effect size relative to other factors.

While there is some evidence to suggest that anonymity leads to toxic disinhibition, some studies have not found this relationship, which indicates that the effect size might be relatively small. Bae [49] found in an experimental research design that anonymity led to a greater feeling of in-group similarity and more attitude change, but less flaming and fewer critical comments. This result seems to directly contradict the other studies mentioned above, but one explanation could be in the topic of conversation and the purpose of the communication. Imagine a meeting, where all are anonymous and dealing with a problem in common. Such a meeting may be conducive of empathy even for complete strangers. Thus, it is not unconceivable that anonymity may enhance empathy between individuals, and recognizing others as more self-similar, especially if other attributes, such as social class, are hidden.

Researchers have suggested other possible explanations for toxic disinhibition. Berg [50] studied the effect of *issue controversy* and found that it had a greater impact on discussion quality than anonymity, suggesting that even if anonymity leads to a decrease in civility and politeness, what is debated (the topic) has a greater effect than if the debaters are anonymous or not. These results are supported by Ksiazek [51] who found that more people commented on certain topics, and that certain topics were more likely to result in uncivil discussions.

Suler [8] and Suler [36] suggested several explanations in addition to anonymity. *Invisibility*, the feeling of not being seen by those one communicates with, regardless of one's anonymity, is thought to be one possible factor contributing to disinhibited behavior. Another suggested contributor, *asynchronicity*, removes the constant feedback-loop of face-to-face communication. *Solipsistic introjection*, when a person reading a message experiences it as a voice within his or her head, can make the sender of the message become a character within one's intrapsychic world. *Dissociative imagination*, which refers to when one has the experience of the created character existing in a different world, may result in online interactions being experienced like a game. *Attenuated status and authority* due to lack of real-world ques of status and authority may also be a factor, especially with regards to how commenters react to moderators. *Perceived privacy* may cause commenters to experience themselves as being in a private encounter online, when they should know better. Finally, *social facilitation*, where the social environment reinforces or fail to counteract disinhibited behavior, is thought to be an important contributing factor to disinhibited behavior.

Suler is not the only researcher to point out the possibility of social influence contributing to negative online behaviors. *Conformity*, defined by Gilovich et al. [30] as the changing of behavior in response to real or imagined, explicit or implicit, pressure from others, is a powerful influencer on behavior, in both positive and negative directions. Participants on online bulletin boards have been found to conform by adopting to both positive and negative information posted by others [52]. These findings are supported by Rösner and Krämer [53], who found that a commenter is more likely to write aggressive comments if peer commenters are aggressive. The frequency of commenting may also be a factor in incivility, as frequent commenters have been found to be less civil and less informal [54]. Although, Coe, Kenski and Rains [55] found the opposite to be true, which indicates that there are other factors at play.

While there are certainly many factors that are thought to be contributing to toxic disinhibition, there is evidence to suggest that anonymity is an important factor. Moreover, anonymity is a popular topic of discussion among researchers, and in the media, when trying to explain toxicity. However, the evidence is inconclusive. Previous studies in experimental settings may not correctly reflect natural conditions. In studies sampling data from online sources, different platforms or populations may influence the results. The current study aims to ameliorate this by sampling comments from similar sources that allow for both anonymous and non-anonymous commenters. However, this is not without problems. We will therefore use random effects to identify sources of variance.

2. Materials and Methods

Two online newspapers were chosen to sample comments from: *The Washington Post* (WP) and *The New York Times* (NYT). These newspapers were chosen because, unlike newspapers that use a Facebook-plugin as a comment section, WP and NYT have comment sections where users must create a separate account. During the account creation they must choose a username, which can either be their real names or a pseudonym. This means that commenters on these platforms make up a population of anonymous and non-anonymous commenters who are all commenting on the same articles on each platform. This reduces the likelihood of interfering variables, such as the affordances of different platforms, with different rules of conduct, moderation and different comment section cultures. Both news sources are east-coast, national, fairly mainstream, left-leaning newspapers [56,57]. Despite apparently using different technologies for moderation, the two newspapers have a similar moderation policy and rules of conduct. Therefore, it is expected that differences in toxic disinhibition can be more stringently and reliably associated with the anonymity of the commenters.

Constructed week sampling was used to create two constructed weeks from February of 2018 to February of 2019 for each newspaper being studied. This involved selecting two random Mondays, two random Tuesdays, etc., during the specified timeframe. This method of sampling is recommended for studying daily newspapers because it creates a randomly selected issue for each day of the week. The events during these days are likely to be referenced in both sources. Two constructed weeks have been found to be sufficient for representing a year's content [58]. In total, 39 articles on politics from the randomly chosen dates were chosen for study. The articles were found using Google's advanced search functions, where one can search for results from a specific website (e.g., nytimes.com), date and subject matter (e.g., politics). There were two requirements for an article to be chosen; (1) the article has to be about politics, and (2) the article must have a substantial number of comments so as to ensure that the data included enough comments from each article to represent the diversity of comments and commenters found in a given comment section. In total, 2451 comments were collected individually and added to a database built for the purpose of securely storing the research data. There were 700 comments were sampled from each newspaper, or 1400 comments in total. During the collection process each comment was coded as being either anonymous or non-anonymous based on their username. From this pool of data, 50 comments were randomly selected for each day and each newspaper, totaling 100 comments for each of the 14 days in the constructed weeks. This adds up to a total number of 1400 of comments sampled for analysis.

The chosen research method for this study was content analysis, which involves establishing categories and counting the number of instances of each category [59]. In this study there would be only two main categories: toxic and neutral. To determine the toxicity of comments they were coded using a coding scheme developed by Papacharissi [10] and used by Rowe [9] was used to categorize the sampled comments. This coding scheme contains 12 categories of uncivil and impolite comments: *threat to democracy, threat to individual rights, stereotypes, name-calling, aspersions, implying disingenuousness, vulgarity, pejorative speak, hyperbole, non-cooperation, sarcasm* and *other* (see appendix for further detail). In the current research, a comment will be labeled as toxic if it fits into any of the 12 categories. In addition to the categories, the coding scheme includes a dimension referred to as *direction.* There are three directions: (1) *Interpersonal* are those comments directed at another commenter; (2) *Other-directed* are comments directed at a specific person or group not present in the comment section; (3) *Neutral* comments are not directed at any specific person or group.

To ensure reliability when determining if a comment was toxic, two coders categorized all 1400 comments. During the coding process, neither coder knew if a comment had been made by an anonymous or non-anonymous commenter, as this information was not presented to the coders during the coding process. After the coders had categorized the comments individually, inter-coder reliability was calculated using Cohen's Kappa a, which is recommended by Hsu and Field [60]. The coders agreed on 91% of the comments, and the inter-coder reliability was found to be 0.73. After the coders had individually coded each comment and inter-coder reliability had been calculated, the two coders met to discuss those comments that they did not agree upon. During this process, the contested comments were discussed, and the coders came to an agreement of which category they both agree on before the final statistical analysis of the data (the detailed instructions to coders are found in Appendix A).

After coding was completed, the data set was analyzed using two methods. An overall association test based on the chi-square test was performed on the coded data to determine if there is a relationship between anonymity and toxic comments. A general linear mixed effects model was developed that used a binomial distribution and a logistic linking function. The formula for the testing involved a linear regression analysis with a dependent variable *toxicity* (yes/no, 1 or 0) being predicted by independent fixed factors anonymity (yes/no), media (NYT/WP) and level (first level or sublevel). All toxic comments were categorized for the direction of the comment either interpersonal (i.e., other commenters), others (including public figures) or neutral (i.e., directed at no particular entity). The model also used commenter identity (a coded signature for anonymous, a coded name for non-anonymous) and the date the comment was written (which is linked to events that happened that day) as random effects to quantify these sources of variance. Random effects assume an open set, i.e., it is assumed that there are many other commenters and many more dates. Fixed effects assume that we deal with a close set, i.e., we are dealing with either anonymity or non-anonymity, either NYT or WP, and it is either first comment or a later comment. This affects how variance is handled by an algorithmic implementation of a general linear model (cf. lme4/glmer, [61]). The results will be presented as odds-ratios, compared to a baseline.

3. Results

Of the 1400 comments, 1181 were written by anonymous commenters and 219 were written by commenters using a real name. When analyzing at all comments from The Washington Post and The New York Times, we see that of the anonymous commenters, 30.7% (n = 363) wrote comments that were coded as toxic. Of the non-anonymous commenters, 20.5% (n = 45) wrote comments that were coded as toxic. In the first analysis, a statistically significant relationship was found between the two variables anonymity and toxic comments ($\chi^2 = 9.3$, $p < 0.002$). The comparison of the count and expected count of anonymous and non-anonymous toxic comments suggests that this relationship is due to non-anonymous commenters being less likely to misbehave in the studied comment sections. Table 1 shows the number of comments for each condition, as well as the expected count if there was no relation between the variables. Analyzing the Washington Post and the New York Times separately produced a similar result, with non-anonymous toxic comments being underrepresented.

Table 1. The count and expected count of anonymous and non-anonymous comments that were coded as toxic and neutral comments.

	Not Anonymous		**Anonymous**	
	Count	Expected	Count	Expected
Toxic	45 (20.5%)	63.8 (29.1%)	363 (30.7%)	344.2 (29.1%)
Neutral	174 (79.5%)	155.2 (70.9%)	818 (69.3%)	836.8 (70.9%)

Below is an Extended Cohen-Friendly graph (cf. [62,63]) that illustrates associations between A toxicity and B anonymity, assuming that all data points are unique examples of comments, but not accounting for writers and dates as sources of variance, which will be analyzed later. The expected number of comments is shown by the dotted lines. The width of each box represents the number of comments in each condition, and their height represents deviation from expected counts. The figure shows that the number of signed toxic comments is significantly lower than expected, suggesting that differences in toxicity between anonymous and non-anonymous commenters could be explained by non-anonymous commenters being less toxic than expected. Significant cells are marked in red. The intent of using association plots is to motivate a more detailed analysis.

As can be seen in Table 1 and Figure 1, there are slightly more toxic comments written by anonymous commenters than expected. However, this is not statistically significant. Rather, for non-anonymous commenters there are significantly fewer toxic comments than expected by chance. In other words, toxic comments among non-anonymous commenters are underrepresented in the data. This variation is statistically significant, indicating that the relational effect is due to non-anonymous commenters behaving better than expected. However, the effect size of the association is tiny (Cramér $\varphi_c = 0.08$).

In the more advanced model, we are able to look closer at sources of variance and we may, therefore, also estimate not only the effect of anonymity, but also the effect of media platform, direction and level of comment, as well as if there is an interaction between anonymity and the strongest other factor.

First, we will examine the data in more detail. Table 2 tabulates comments into neutral and toxic comments for the two websites, divided up by *original and downstream* comments. A comment is labeled an original comment if it comments directly at the article (first position in a thread) and downstream if it is a later comment on a previous comment (adding, following up, simply staying within the thread). We see that comments in Washington Post generally have a higher proportion of toxic comments. This difference between NYT and WP has a small effect size of $\varphi_c = 0.16$ (signed) and $\varphi_c = 0.15$ (anonymous). However, being downstream does not alter the proportion of toxic comments. This will be investigated further using a statistical model.

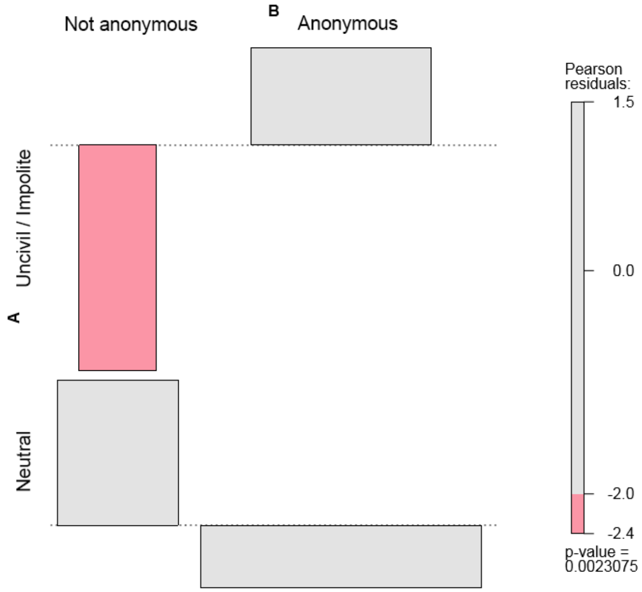


Figure 1. Cohen-Friendly graph of Table 1. Less toxic for signed comments.

Table 2. Proportions of toxic comments in New York Times and Washington Post. For both journals, anonymity is associated with more toxic comments.

	NYT		WP	
	Original	**Downstream**	**Original**	**Downstream**
Signed comments				
neutral	82% (51)	84% (86)	77% (17)	61% (20)
toxic	18% (11)	16% (16)	23% (5)	39% (13)
Anonymous comments				
neutral	74% (142)	79% (271)	54% (143)	69% (262)
toxic	26% (50)	21% (73)	46% (120)	31% (120)

The model is based on a general linear mixed effects model [61] fit by maximum likelihood and using a binomial distribution with a logistic linking function. Commenters may contribute more than one data point, and there are many data points for each date. This will be handled by random effects assigned for identification codes for the commenters and the dates. The Mixed Effect design treats them as sources of variance and may handle these sources simultaneously (more details in Appendix B).

$$\text{Toxic} \sim \text{Anonymity} * \text{Website} + \text{Level} + (1 \mid \text{Date}) + (1 \mid \text{Id}) \tag{1}$$

Formula (1) simply states that we try to explain toxicity in terms of (a) anonymity possibly interacting with website (b) (Response) Level (original/downstream). These are our fixed effects. We are further modeling the sources of variance stemming from (a) the date the comment was written, and (b) the identifier of the commenter. These are our random effects used to control the variance stemming from individuals (id) and events (days).

Caveats: We cannot know if there is only one person behind each identifier. It is possible that more than one person may share a signature, or that an account has been accessed by an unauthorized person. In total we have 1400 data points, and 1083 individuals were identified (Id) in 15 different days (Date). There are relatively few different dates sampled. However, the dates are considered fairly average dates with no extraordinary events.

Figure 2 gives the odds ratios of our fixed factors. Anonymous is *not* significant (z = 1.407 p = 0.141) with 45% more toxic comments (1.45). Website is significant (z = 2.460 p = 0.014) and Washington Post is associated with about 2.61 times the rate of toxic comments. Response level is significant and tend towards *less* toxic comments (z = −2.357 p = 0.018) for downstream comments. We did not detect a significant interaction between anonymity and website.

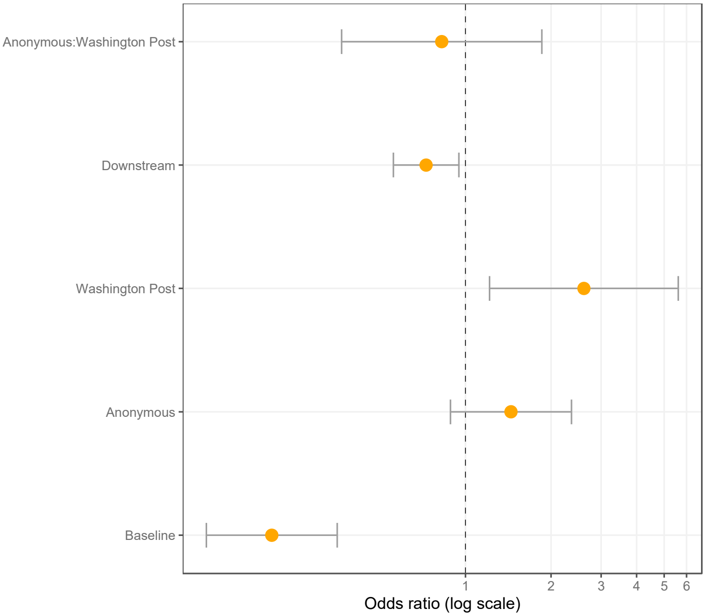


Figure 2. Odds ratio for the fixed factors. Baseline is New York Times, signed, original comments. From Table 3 the baseline is about 18% toxic comments, which is congruent with model estimates ((20.8 ± 1.3)%). Odds ratio of 1 means no change.

Table 3. Number of toxic comments coded as interpersonal, other-directed and neutral.

Interpersonal			**Other-Directed**			**Neutral**		
WP	**NYT**	**Total**	**WP**	**NYT**	**total**	**WP**	**NYT**	**Total**
117	58	175	129	78	207	12	14	26

Table 3 shows that 175 of the toxic comments were interpersonal and directed at other commenters, 207 were other-directed, meaning they were directed at persons or groups not present in the comment section, and only 26 comments were neutral, meaning that they were not directed at any specific person or group.

During the coding process, comments directed at public figures, such as politicians, were specifically marked as being directed at a public figure, in addition to being coded as *other-directed* (Table 4, Figure 3). This subcategory was added to further explore other-directed toxicity. In total, 115 of the 207 other-directed comments were directed specifically towards public figures. While toxic comments directed towards public figures are problematic, there is an argument to be made that the way one speaks about public figures is not the same as when speaking of, for example, other commenters or private individuals. Therefore, we argue that in future research using this coding scheme, the category of *other-directed* could be further divided into two categories; comments directed at public figures and comments directed at private individuals.

Table 4. Proportions of toxic comments divided up on what the comments are directed at: 1 is interpersonal 2 is directed at other (including public figures) and 3 is neutral.

	NYT			WP		
	1	**2**	**3**	**1**	**2**	**3**
Signed comments	12% (7)	24% (19)	7% (1)	8% (9)	6% (8)	8% (1)
Anonymous comments	88% (51)	76% (59)	93% (13)	92% (108)	94% (121)	92% (11)

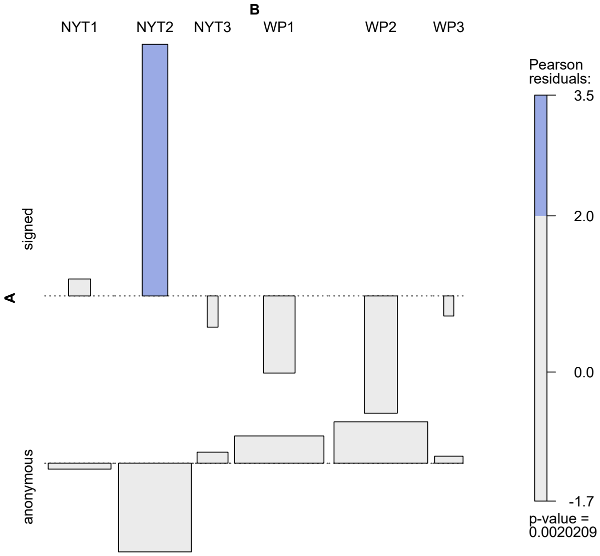


Figure 3. The Cohen-Friendly graph of Table 4 shows that toxic comments directed at others (including public figures) are more associated with signed comments ($p < 0.01$) in the New York Times.

4. Discussion

Comment sections on news sites have the potential to serve as an important channel for public debate. They allow people to express themselves on a variety of topics, with a large potential audience that includes the journalists who wrote the articles being commented on. However, if comment sections are to be a welcoming forum of expression for everyone, it is important to understand why some commenters choose to write toxic and derogatory comments. Our study attempts a contribution to ongoing research on the role of anonymity, but by using a sampling strategy that we believe will provide more accurate results.

We found statistically significant relationships that contribute to understanding toxic disinhibition in the comment sections of The Washington Post and The New York Times. Both can be described as left-leaning media. It is an interesting extension to investigate the effect of political association, on a scale from left to right, but this demands a much larger study. We have decided not to use this dimension, and one motivation is that the political association of the commenters is still unknown.

The result of this study suggests that there is an association between anonymity and toxic comments. Non-anonymous commenters wrote fewer toxic comments than is expected if all were equal. We interpret this to mean that anonymity may have an effect on toxicity, but it is the lack of anonymity that makes a commenter less toxic. In other words, anonymity does not cause toxic comments, but signing a comment either makes commenters behave better, or possibly signed contributors are associated with more proficient writers.

We found out that there are stronger differences between the two platforms. While anonymity may affect toxicity, editing policies play an equally important role. Because the Washington Post is associated with anonymous toxic comments that website is a stronger explanation for toxicity than anonymity alone. The New York Times may be more active in enforcing their rules of conduct and thus more toxic comments may have been deleted there. The Washington Post and The New York Times have extensive community rules and guidelines that are linked to in the comment sections [64,65]. The rules of conduct themselves give no indication why there would be a difference in toxicity between the two newspapers. While the Washington Post's guidelines are more extensive, both newspapers have guidelines that reflect their desire for civil and well-informed comments, and neither allow personal attacks, vulgarity or off-topic comments. The differences between the two newspapers could be explained by differences in moderation. We do not know how many moderators each newspaper employs, how they work and by what standards they moderate. We do know that the New York Times uses a semi-automated system for effective moderation. In partnership with the Alphabet-owned company Jigsaw, they use machine learning technology for moderation, allowing them to keep comment sections open longer without overextending the resources spent on moderation [66]. It is possible that this system is better at catching unwanted comments than the system used by The Washington Post. The issue of automatic moderation is complicated by the complexity of the task and creative use of language. The state-of-the-art technology for the related task of sentiment detection shows a combined measure of precision and recall between 0.60 and 0.89 (and similar ranges for accuracy) for a wide range of algorithms used on controlled datasets on product and hotel reviews [67]. Chen et al. [68] used Convolutional Neural Networks, with some preprocessing, to detect verbal aggression in Twitter comments with similar results on their test sets. Their test accuracy reached at most about 90% [68]: Figures 7 and 9. Xu et al. [69] show similar results on sentiment detection in comment fields. Algorithms tend to behave worse on truly novel texts outside of the training data, but more data and continuously retraining models may compensate. Even with access to very large databases and deep learning algorithms, there is thus room for either missing a sentiment or mislabeling. In the case of automatic moderation of toxicity, it may create frustration for users if their comments are erroneously publicly flagged or edited out.

It should be noted that the findings in the present study are fairly robust, and the effect of anonymity was detected by different methods. Models that excluded interaction between website and anonymity, and excluded response level, were also tested. The results were very similar. The reason for giving the more elaborate model is to show that other available factors were not responsible for the results. There might, however, be other factors that were not available or controlled in our study.

While the observed relationship between anonymity and toxic comments is interesting, it is important to acknowledge that other associations are stronger, making it difficult to conclude with certainty that anonymity is a significant cause for toxic disinhibition in comment sections. Previous research has concluded that anonymity leads to greater toxicity in comment sections [9,11,44,45]. As mentioned previously, these studies sample data from multiple sources, which could potentially lead to results being skewed by uncontrolled variables. The current study sampled anonymous and non-anonymous comments from the same platforms. While we did find an association between anonymity and toxicity, the result of this study suggests that anonymity has a small effect on the civility of online comment sections. While anonymity may affect toxicity in comment sections, it is certainly not the only factor that should be considered. Issue controversy may play an important role in how commenters debate, as Berg [50] suggested. The comments analyzed for this study were written on political articles at a time of much political controversy and in a highly polarized political climate. However, both anonymous and non-anonymous commenters should be equally affected by political tensions and issue controversy, assuming that people have honest intentions to discuss the issues. A competing hypothesis is that people choose to be anonymous when they have malicious intent, i.e., intend to disrupt a conversation. This is not supported by our data.

Social influence is another important aspect that should be considered. As stated earlier, conformity has been found to effect toxicity in online communication. It is possible that anonymous and non-anonymous commenters are affected differently by social influence. As noted earlier, anonymity has been found to lead to a greater feeling of in-group similarity and more attitude change [49]. If being anonymous affects a commenter's feelings of similarity to other commenters this could certainly be thought to affect the toxicity of anonymous comments. If we can encourage writers to stay on topic and show more compassion with people or views they do not agree with, then we may ameliorate the negative effects of anonymity, without policing language or opinions.

While the results of this study are interesting, it is important to be aware of its limitations. Firstly, we sampled comments from just two newspapers within a limited time period. Different newspapers use different technological solutions to facilitate commenting, which through affordances, design and moderation policies could be thought to influence the discussions among commenters. Indeed, we detected a significant difference between our two very similar platforms. However, the effect might be platform internal or external. One internal explanation is that platforms, despite having similar rules of conduct, have different editing policies. An external explanation is that the population of commenters may be different between platforms or between levels of anonymity. There may well be larger differences between populations between other platforms, as we chose the examined platforms for their apparently similar political and geographical appeal.

Newspapers use moderators to check for and delete comments that are against the rules of conduct or require deletion for legal reasons. It is possible that comments have been deleted before they could be sampled for this study, and the inclusion of these deleted comments may have had an effect on the results. Therefore, it is accurate to say that our results are limited by an apparent survivor bias.

The comment sections of both The Washington Post and The New York Times allow for users to create any username, and it is possible that some commenters have created pseudonyms that appear to be real names. Obviously fake names, such as *Darth Vader,* were coded as being anonymous during the sampling process. It was not possible for us to verify the identity of commenters using a real-looking name. The websites have more information available; however, sharing such information with a third party violates privacy.

A commenter that wanted to use a pseudonym, would most likely create an obvious pseudonym and not a real-looking name, unless they are sailing under a false flag, which violates standard agreements for setting up a user account. On a platform that allows for pseudonyms, especially one where pseudonyms are the norms, there is little reason for someone to create a name that appears to be a real one.

5. Conclusions

The current study has attempted to improve on the methodology of researching anonymity's effect on toxicity by sampling data from comment sections where anonymous and non-anonymous users debate on the same platform. This novel sampling strategy makes us confident in the results of the statistical tests.

We have found a small but significant relationship between anonymity and toxic comments. At first sight this result seems to support the prevailing view that anonymity causes toxic behavior. However, the data suggest that it is non-anonymous users who are less toxic than expected, and not anonymous users being more toxic. A simpler explanation could be that signed writers are more proficient writers. The effect size of anonymity is tiny or small. Our own analysis showed that the effects of platform and the direction of the comment were stronger than the effect of anonymity. Another interesting finding is the fact that non-anonymous comments were *less* toxic than expected, while anonymous comments were not significantly more toxic than expected. This is congruent with the observed effect of durable pseudonymity [47], where the quality of comments improved over time for durable pseudonyms. Many anonymous commentors may choose anonymity, not to troll others but to avoid personal attacks in real life. Thus, anonymity is valuable for a freer more democratic debate, and the quality of debate may be improved by fairly simple measures, such as encouraging durable pseudonymity.

Previous research has found other explanations for online toxicity, such as issue controversy [50,51] and social influence [36,52,53]. In our opinion, it is important to evaluate the causes of problematic online behavior. One controllable factor, apart from simply editing out toxic comments (or commenters), is to enforce a discussion to stay on topic and not comment on other users or public figures. As discussed, there are also many positive aspects of anonymity that are at risk if anonymity is cancelled. The small reduction in toxicity may negatively affect the expected quality of comments and limit the diversity of opinions.

Author Contributions: Conceptualization, M.K.; methodology, M.K.; formal analysis, C.J.; data curation, M.K.; writing—original draft preparation, M.K. and C.J.; writing—review and editing, M.K. and C.J.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. Study involve investigating comments made in the public domain. Commentors are not identified by neither name nor signature in this article.

Informed Consent Statement: In accordance with the Norwegian Centre for Research Data, who approved the methodology of this study, consent was waived due to the low impact on research subjects, the public nature of the collected data, and the impracticality of gathering consent from already anonymous subjects.

Data Availability Statement: In accordance with the Norwegian Centre for Research Data, who approved the methodology of this study, data is not available.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Code "1" all comments containing a "threat to democracy": A comment ought to be coded as containing a threat to democracy if it advocates the overthrow of the government (i.e., if it proposes a revolution) or if it advocates an armed struggle in opposition to the government (i.e., if the commenter threatens the use of violence against the government). Examples of such threats include commenters suggesting that government efforts to restrict guns, for example, would lead them to take up arms. For example, one commenter suggested that if the government were to enforce the ban on assault weapons and try and take his gun, "they would soon regret it". Similarly, commenters threatening to start a revolution in response to the government implementing policy would also be coded as a threat to democracy.

Exceptions: Should you believe that the threat is sarcastic, please code for 'sarcasm' (11), not a threat to democracy. "Non-cooperation" (8) should also not be confused with a threat to democracy.

Code "2" all comments containing a "threat to individual rights": A comment ought to be coded as containing a threat to individual rights if it advocates restricting the rights or freedoms of certain members of society or certain individuals. Such examples are common when sensitive or divisive political issues are being discussed because commenters often resort to threatening one another or often advocate restricting the rights of groups or individuals they blame for the event which led the issue to being discussed. For example, following a tragic shooting in which a psychologically disturbed individual is implicated, many people are quick to suggest that the rights of mentally ill citizens be restricted, e.g., "They should all be locked up" would be an example of this. Furthermore, supporters of gun-control often blame those who oppose gun-control, for example, for the widespread use of guns and, by extension, such tragic events. In doing so, they suggest that it is they who are responsible for such tragedies and, therefore, "they have no right to participate in this debate." Exceptions: Threats to individual rights should not be confused with stereotypes (although they might be closely related if the threat being made assumes that all members of that particular group is the same) or with non-cooperation. Refusing to co-operate is not necessarily the same as refusing others the right to participate in the discussion.

Code "3" all comments containing the use of "stereotypes": A comment ought to be coded as containing a stereotype if it asserts a widely held but fixed and oversimplified image or idea of a particular type of person or thing. This includes associating people with a group using labels, whether those are mild—"liberal", or more offensive—"faggot". The use of stereotypes is common when the topic being discussed is highly partisan.

Stereotyping may also involve making generalized assumptions about the thoughts and behavior of certain groups or individuals based on said stereotypes, for example, suggesting gun-owners/supporters are paranoid, liberals/conservatives are less/more patriotic, or immigrants rely heavily upon social security.

Exceptions: The use of the words liberal or conservative are not always used stereotypically. For example, an administration or an individual may be liberal or conservative in their views, but this type of description is not necessarily stereotypical or derisory.

Note: Stereotypes should also be coded for their direction: those intended to offend others should be coded as antagonistic (e.g., "you liberals are all the same. You want to ban anything you don't like and that doesn't suit you.") or neutral if it was used in articulating an argument but without the intent to offend others (e.g., "the liberal agenda has caused a huge rise in regulations across a number of industries").

Code "4" all comments containing "name-calling": (e.g., gun-nut, idiot, fool, etc.). To be coded as name-calling the words used must be clearly derogatory towards the person it is intended for. Exceptions: Be careful not to include words which may be regarded as a stereotype (e.g., liberal). If name-calling is aimed at a group, or the "name" is often applied to a group of individuals, it may potentially be a stereotypical comment (e.g., anyone who owns a gun is an idiot—this groups all gun-owners together, therefore stereotyping them).

Code "5" all comments containing "aspersions": All comments containing "an attack on the reputation or integrity of someone or something" ought to be coded for aspersion. A comment may be coded as including an aspersion if it contains disparaging or belittling comments aimed at other commenters or their ideas. These ought to include explicit efforts to express dismay at others. For example, a comment which reads: "Teachers don't need to be carrying guns! It's stupid!" may be considered an aspersion. A comment which reads: "sheer idiocy" may also be considered an aspersion. Similarly, a comment which reads: "this is a free country that prohibits slavery. Do you have a problem with that?" may also be coded as an aspersion as its tone implies it is not a genuine question, but an attack on a previous comment/idea. An aspersion may be both explicit or implicit.

Code "6" all comments containing "lying": All comments implying disingenuousness (e.g., liar, dishonest, fraud etc.) of other commenters or public figures ought to be coded as lying Exceptions: If a comment casts doubt on the truthfulness of a previous comment or a public figure this does not constitute the use of synonyms for liar. For example, if a commenter writes "that is not true", they are not implying that the other person is intentionally lying, but rather that they are misinformed.

Code "7" all comments containing vulgarity: All comments containing vulgar language (e.g., crap, shit, any swear-words/cursing, sexual innuendo etc.) ought to be coded as vulgar. Comments containing vulgar abbreviations such as WTF (what the fuck) should also be coded as vulgar.

Code "8" all comments containing "pejorative speak": All comments containing language which disparages the manner in which someone communicates (e.g., blather, crying, moaning, etc.) ought to be coded as pejorative for speech.

Code "9" all comments containing "hyperbole": Comments which contain a massive overstatement (e.g., makes pulling teeth with pliers look easy) ought to be coded as hyperbole. Be careful not to include words which accurately describe events, particularly given that many of the topics under discussion may be described using words associated with hyperbole (e.g., the Newtown shooting may be described both as a "massacre" and a "heinous" act), although these words are not necessarily used to overemphasize it. Hyperbole might be characterized either as a phrase (e.g., barely a week goes by without a shooting), or the overuse of descriptive words designed to emphasize a point (e.g., "It's not the guns that kill but a ticking time bomb of anger seething in society, giving clues & everyone ignoring him until he kills little babies with an illegal automatic weapon. I don't think it was an accident he killed mommy, the Ph.D. & Principal. He was suicidal & homicidal; very common & wanted notoriety. What better way than to kill babies"). Note: many social issues are discussed using language which may be considered hyperbole, e.g., abortion = murder, gay marriage = abomination, etc. It is up to you as to whether you believe the commenter is making an overstatement or just describes it as such.

Code "10" all comments containing "non-cooperation": The discussion of a situation in terms of a stalemate ought to be coded as non-cooperation. Outright rejection of an idea/policy by a commenter should only count as non-cooperation if it involves excessive use of exclamation marks or capital letters for example. For example, a comment which reads: "I'm 48 years old. I retired after 20 years in the military. I went back to college to be a special education teacher. I WILL NEVER CARRY A FIREARM INTO MY CLASSROOM." Find another solution' may be considered non-cooperation. Similarly, a comment which reads: "I hate guns!! I refuse to send my kids to a school where the teachers are armed!!!!!!!" may be coded as non-cooperation.

Exceptions: A simple rejection of an idea/policy should not be considered non-cooperation. Likewise, suggesting that another commenter has no right to take part in the discussion for whatever reason should be coded as "threat to individual rights" insofar as it threatens their right to free speech, not as non-cooperation. Only a refusal to listen or comply should be coded as non-cooperation.

Code "11" all comments containing "sarcasm": "You'll know it when you see it!!"

Code "12" all comments which may be deemed impolite, but which do not fall into any of the previous categories of impoliteness: This category ought to catch any other type of impoliteness that you think is evident and which does not fit into any other category above. This most commonly includes using capital letters to symbolize shouting and the use of blasphemous language. Even comments you believe are impolite in their tone may be coded as "other" (12).

Exceptions: CAPITAL LETTERS, if used for single words, should be assumed to be signaling emphasis. If a phrase or sentence is written in CAPS, this may be considered shouting.

Direction of Incivility:

All uncivil and impolite comments should be coded for their direction, with the exception of stereotypes which should be coded as antagonistic or neutral. Once the type of incivility has been categorized, the direction then needs to be coded. Comments containing incivility and which are aimed at another commenter in the discussion should be coded as Interpersonal (i). Interpersonal comments include those which are explicitly directed at other commenters (e.g., where the comment includes the name of other commenters) or those which address the comments of others, even without naming them. An example of interpersonal incivility may include: "I can't wait to see you on the battlefield someday Leo [another commenter] because that is what it's gonna boil down to . . . .you believe what you want and you should BUT DO NOT FORCE YOUR BELIEFS ON ME". If the comment contains incivility and is aimed at a specific person or group of people not present, the comment is coded as Other-directed (od). In this case, the "other" often refers to a politician (e.g., Obama), a pressure group (e.g., the NRA), a political party (e.g., Republicans), the media (e.g., the Washington Post) or state institutions (e.g., SCOTUS). If the comment contains incivility but does not refer, or imply reference, to another commenter or 'other', the comment is coded as Neutral (n). Neutral incivility occurs primarily when the commenter disagrees with the content of the article being commented on. An example of neutral incivility may include: "A Bushmaster in a classroom? WTF!!" The direction of a comment is very much dependent on the coders' understanding of whether or not it refers to other comments in the thread or whether it is a stand-alone comment which is not intended as a response. Thus, it is important to be familiar with the content and language of the article to which the comment refers.

Appendix B

Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) [glmerMod]
Family: binomial (logit)
Formula:
Toxic ~ Anonymity * Website + R + (1 | Date) + (1 | Id)
Data: magnus
Control: glmerControl(optimizer = "bobyqa")

AIC	BIC	logLik	Deviance	df.resid
1617.9	1654.6	−802.0	1603.9	1393

Table A1. Scaled residuals.

Min	1Q	Median	3Q	Max
−1.2282	−0.6312	−0.4597	0.9508	2.6189

Table A2. Random effects.

Groups	Name	Variance	Std.Dev.
Id	(Intercept)	0.3259	0.5708
Date	(Intercept)	0.2078	0.4559

Number of obs: 1400, groups: Id, 1083; Date, 15.

Table A3. Fixed effects.

	Estimate	Std. Error	z	Pr(> |z|)
(Intercept)	−1.5694	0.2705	−5.803	6.53×10^{-9} ***
Anonymity=1(anonymous)	0.3685	0.2502	1.473	0.1407 (n.s.)
Website = washingtonpost	0.9595	0.3900	2.460	0.0139 *
Response Level = 2	−0.3193	0.1355	−2.357	0.0184 *
Anonymity = 1: Website = Washington Post	−0.1924	0.4138	−0.465	0.6420 (n.s)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table A4. Correlation of Fixed Effects.

	Intercept	Anonymous	Washington Post	Downstream
Anonymous	−0.730			
Washington Post	−0.496	0.515		
Downstream	−0.273	−0.023	0.002	
Anonymous and WP	0.442	−0.606	−0.931	0.005

References

1. Finley, K. A Brief History of the End of the Comments. Available online: https://www.wired.com/2015/10/brief-history-of-the-demise-of-the-comments-timeline/ (accessed on 2 March 2021).
2. Wallsten, K.; Tarsi, M. Persuasion from Below? *J. Pract.* **2015**, *10*, 1019–1040. [CrossRef]
3. Bilton, R. Why Some Publishers Are Killing Their Comment Sections. *Digiday UK* **2014**, *14*. Available online: https://digiday.com/media/comments-sections/ (accessed on 2 March 2021).
4. Ellis, J. *What Happened after 7 News Sites Got Rid of Reader Comments*; Neiman Lab.: Cambridge, MA, USA, 2015.
5. Ramnefjell, G. Dagbladets Kommentarfelt (1996–2016); Dagbladet.no. 2016. Available online: https://www.dagbladet.no/kultur/dagbladets-kommentarfelt-1996---2016/60160514 (accessed on 2 March 2021).
6. Waatland, E. Nettavisen Stenger Kommentarfeltet Med Umiddelbar Virkning. [The Net-Journal Closes Their Comment Field Effective Immediately]. Available online: https://m24.no/erik-stephansen-gunnar-stavrum-kommentarfelt/nettavisen-stenger-kommentarfeltet-med-umiddelbar-virkning/205456 (accessed on 2 March 2021).
7. Reagle, J.M. *Reading the Comments: Likers, Haters, and Manipulators at the Bottom of the Web*; MIT Press: Sabon, NY, USA, 2015.
8. Suler, J. The Online Disinhibition Effect. *Int. J. Appl. Psychoanal. Stud.* **2005**, *2*, 184–188. [CrossRef]
9. Rowe, I. Civility 2.0: A comparative analysis of incivility in online political discussion. *Inf. Commun. Soc.* **2014**, *18*, 121–138. [CrossRef]
10. Papacharissi, Z. Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media Soc.* **2004**, *6*, 259–283. [CrossRef]
11. Santana, A.D. Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *J. Pract.* **2014**, *8*, 18–33. [CrossRef]
12. Lapidot-Lefler, N.; Barak, A. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Comput. Hum. Behav.* **2012**, *28*, 434–443. [CrossRef]
13. Zimmerman, A.G.; Ybarra, G.J. Online aggression: The influences of anonymity and social modeling. *Psychol. Pop. Media Cult.* **2016**, *5*, 181–193. [CrossRef]
14. Hubler, M.T.; Bell, D.C. Computer-mediated humor and ethos: Exploring threads of constitutive laughter in online communities. *Comput. Compos.* **2003**, *20*, 277–294. [CrossRef]
15. Gonçalves, J. A peaceful pyramid? Hierarchy and anonymity in newspaper comment sections. *Observatorio* **2015**, *9*, 1–13.
16. Li, X. *Internet Newspapers: The Making of a Mainstream Medium*; Routledge: New York, NY, USA, 2010.
17. Løvlie, A.S. Constructive Comments?: Designing an online debate system for the Danish Broadcasting Corporation. *J. Pract.* **2018**, *12*, 781–798. [CrossRef]
18. Singer, J.B.; Paulussen, S.; Hermida, A. *Participatory Journalism: Guarding Open Gates at Online Newspapers*; Wiley-Blackwell: Malden, MA, USA, 2011.
19. Stroud, N.J.; Muddiman, A.; Scacco, J.M. Like, recommend, or respect? Altering political behavior in news comment sections. *New Media Soc.* **2016**, *19*, 1–17. [CrossRef]
20. Artime, M. Angry and Alone: Demographic Characteristics of Those Who Post to Online Comment Sections. *Soc. Sci.* **2016**, *5*, 68. [CrossRef]
21. Toepfl, F.; Piwoni, E. Public Spheres in Interaction: Comment Sections of News Websites as Counterpublic Spaces. *J. Commun.* **2015**, *65*, 465–488. [CrossRef]
22. Graf, J.; Erba, J.; Harn, R.-W. The Role of Civility and Anonymity on Perceptions of Online Comments. *Mass Commun. Soc.* **2017**, *20*, 526–549. [CrossRef]
23. Vergeer, M. Twitter and Political Campaigning. *Sociol. Compass* **2015**, *9*, 745–760. [CrossRef]
24. Ksiazek, T.B. Civil Interactivity: How News Organizations' Commenting Policies Explain Civility and Hostility in User Comments. *J. Broadcast. Electron. Media* **2015**, *59*, 556–573. [CrossRef]
25. Scott, C.R. Benefits and Drawbacks of Anonymous Online Communication: Legal Challenges and Communicative Recommendations. *Free Speech Yearb.* **2012**, *41*, 127–141. [CrossRef]
26. Hogan, B. Pseudonyms and the Rise of the Real-Name Web. In *A Companion to New Media Dynamics*; Hartley, J., Burgess, J., Bruns, A., Eds.; Blackwell publishing Ltd.: Chichester, UK, 2013; pp. 290–308.
27. Savill, R. Harry Potter and the Mystery of J K's Lost Initial. Available online: https://www.telegraph.co.uk/news/uknews/1349288/Harry-Potter-and-the-mystery-of-J-Ks-lost-initial.html (accessed on 2 March 2021).
28. LeBon, G. The crowd: A study of the popular mind. In *Crowd*; T.F. Unwin: London, UK, 1908.
29. Minot, C.S. The Crowd: A Study of the Popular Mind. *Psychol. Rev.* **1897**, *4*, 313–316.
30. Gilovich, T.; Keltner, D.; Chen, S.; Nisbett, R.E. *Social Psychology*; W.W. Norton & Company Ltd.: London, UK, 2016.
31. Festinger, L.; Pepitone, A.; Newcomb, T. Some consequences of de-individuation in a group. *J. Abnorm. Soc. Psychol.* **1952**, *47*, 382–389. [CrossRef]
32. Diener, E.; Fraser, S.C.; Beaman, A.L.; Kelem, R.T. Effects of deindividuation variables on stealing among Halloween trick-or-treaters. *J. Personal. Soc. Psychol.* **1976**, *33*, 178–183. [CrossRef]
33. Felipe, V.; Beria, F.M.; Costa, Â.B.; Koller, S.H. Deindividuation: From Le Bon to the Social Identity Model of Deindividuation Effects. Available online: https://psycnet.apa.org/record/2017-56729-001 (accessed on 2 March 2021).

34. Reicher, S.D.; Spears, R.; Postmes, T. A Social Identity Model of Deindividuation Phenomena. *Eur. Rev. Soc. Psychol.* **1995**, *6*, 161–198. [CrossRef]
35. Kirkpatrick, D. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*; Simon & Schuster Paperbacks: New York, NY, USA, 2011.
36. Suler, J. *Psychology of the Digital Age: Humans Become Electric*; Cambridge University Press: New York, NY, USA, 2016.
37. Marwick, A.E.; Boyd, D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media Soc.* **2010**, *13*, 114–133. [CrossRef]
38. Rosenberry, J. Users Support Online Anonymity despite Increasing Negativity. *Newsp. Res. J.* **2011**, *32*, 6–19. [CrossRef]
39. Jacobsen, C.; Fosgaard, T.R.; Pascual-Ezama, D. Why Do We Lie? A Practical Guide to the Dishonesty Literature. *J. Econ. Surv.* **2018**, *32*, 357–387. [CrossRef]
40. Stein, E. Queers anonymous: Lesbians, gay men, free speech, and cyberspace. *Harv. Civ. Rights Civ. Liberties Law Rev.* **2003**, *38*, 159–213. [CrossRef]
41. Hirsh, J.B.; Galinsky, A.D.; Zhong, C.B. Drunk, Powerful, and in the Dark: How General Processes of Disinhibition Produce Both Prosocial and Antisocial Behavior. *Perspect. Psychol. Sci.* **2011**, *6*, 415–427. [CrossRef]
42. Postmes, T.; Spears, R.; Sakhel, K.; de Groot, D. Social Influence in Computer-Mediated Communication: The Effects of Anonymity on Group Behavior. *Personal. Soc. Psychol. Bull.* **2001**, *27*, 1243–1254. [CrossRef]
43. Rowe, I. Deliberation 2.0: Comparing the Deliberative Quality of Online News User Comments Across Platforms. *J. Broadcast. Electron. Media* **2015**, *59*, 539–555. [CrossRef]
44. Dillon, K.P.; Neo, R.L.; Seely, N. Civil keystrokes: Examining anonymity, politeness, and civility in online newspaper forums. In *Internet Research 16*; The 16th Annual Meeting of the Association of Internet Researchers: Pheonix, AZ, USA, 2015.
45. Santana, A.D. Toward quality discourse: Measuring the effect of user identity in commenting forums. *Newsp. Res. J.* **2019**, *40*, 467–486. [CrossRef]
46. Fredheim, R.; Moore, A.; Naughton, J. Anonymity and Online Commenting: The Broken Windows Effect and the End of Drive-by Commenting. Available online: https://dl.acm.org/doi/abs/10.1145/2786451.2786459 (accessed on 2 March 2021).
47. Moore, A.J.; Fredheim, R.; Wyss, D.; Beste, S. Deliberation and Identity Rules: The Effect of Anonymity, Pseudonyms and Real-Name Requirements on the Cognitive Complexity of Online News Comments. *Political Stud.* **2021**, *69*, 45–65. [CrossRef]
48. Barlett, C.P.; Gentile, D.A.; Chew, C. Predicting cyberbullying from anonymity. *Psychol. Pop. Media Cult.* **2016**, *5*, 171–180. [CrossRef]
49. Bae, M. The effects of anonymity on computer-mediated communication: The case of independent versus interdependent self-construal influence. *Comput. Hum. Behav.* **2016**, *55*, 300–309. [CrossRef]
50. Berg, J. The impact of anonymity and issue controversiality on the quality of online discussion. *J. Inf. Technol. Politics* **2016**, *13*, 37–51. [CrossRef]
51. Ksiazek, T.B. Commenting on the News: Explaining the degree and quality of user comments on news websites. *J. Stud.* **2018**, *19*, 650–673. [CrossRef]
52. Cheng, S.L.; Lin, W.-H.; Phoa, F.K.H.; Hwang, J.-S.; Liu, W.-C. Analysing the Unequal Effects of Positive and Negative Information on the Behavior of Users of a Taiwanese On-Line Bulletin Board. *PLoS ONE* **2015**, *10*, e0137842.
53. Rösner, L.; Krämer, N.C. Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Soc. Media Soc.* **2016**, *16*, 1–13. [CrossRef]
54. Blom, R.; Carpenter, S.; Bowe, B.J.; Lange, R. Frequent Contributors Within U.S. Newspaper Comment Forums: An Examination of Their Civility and Information Value. *Am. Behav. Sci.* **2014**, *58*, 1314–1328. [CrossRef]
55. Coe, K.; Kenski, K.; Rains, S.A. Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *J. Commun.* **2014**, *64*, 658–679. [CrossRef]
56. Media Bias Ratings. 2019. Available online: https://www.allsides.com/media-bias/media-bias-ratings?field_featured_bias_rating_value=All&field_news_source_type_tid[1]=1&field_news_source_type_tid[2]=2&field_news_source_type_tid[3]=3 (accessed on 18 November 2019).
57. Where Do News Sources fall On the Political Bias Spectrum? 20 December 2018. Available online: https://guides.lib.umich.edu/c.php?g=637508&p=4462444 (accessed on 18 November 2019).
58. Riffe, D.; Lacy, S.; Fico, F. *Analyzing Media Messages: Using Quantitative Content Analysis in Research*; Routledge: New York, NY, USA, 2014.
59. Silverman, D. *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*; Cromwell Press: Townbridge, UK, 2001.
60. Hsu, L.M.; Field, R. Interrater Agreement Measures: Comments on Kappan, Cohen's Kappa, Scott's π, and Aickin's α. *Underst. Stat.* **2010**, *2*, 204–219. [CrossRef]
61. Bates, D.; Machier, M.; Bolker, B.; Walker, S. Fitting Linear Mixed Effects Models using Ime4. *J. Stat. Softw.* **2015**, *67*, 1. [CrossRef]
62. Cohen, A. On the graphical display of the significant components in a two-way contingency table. *Commun. Stat. Theory Methods* **1980**, *A9*, 1025–1041. [CrossRef]
63. Meyer, D.; Zeileis, A.; Hornik, K. Visualizing independence using extended association plots. In Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria, 20–22 March 2003.
64. Amenabar, T. Community Rules. 11 June 2018. Available online: https://www.washingtonpost.com/news/ask-the-post/wp/2018/06/11/community-rules/ (accessed on 7 February 2020).

65. Comments. 2020. Available online: https://help.nytimes.com/hc/en-us/articles/115014792387-Comments (accessed on 7 February 2020).
66. Etim, B. The Times Sharply Increases Articles Open for Comments, Using Google's Technology. 13 June 2017. Available online: https://www.nytimes.com/2017/06/13/insider/have-a-comment-leave-a-comment.html (accessed on 7 February 2020).
67. Dashtipour, K.; Gogate, M.; Li, J.; Jiang, F.; Kong, B.; Hussain, A. A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing* **2020**, *380*, 1–10. [CrossRef]
68. Chen, J.; Yan, S.; Wong, K.-C. Verbal aggression detection on Twitter comments: Convolutional neural network for short-text sentiment analysis. *Neural Comput. Appl.* **2018**, *32*, 1–10. [CrossRef]
69. Xu, G.; Meng, Y.; Qiu, X.; Yu, Z.; Wu, X. Sentiment analysis of comment texts based on BiLSTM. *IEEE Access* **2019**, *7*, 51522–51532. [CrossRef]

PEER-REVIEWED JOURNAL ON THE INTERNET

Get lost, troll: How accusations of trolling in newspaper comment sections affect the debate

by Magnus Knustad

Abstract

This qualitative study explores instances where someone is accused of being a troll or a bot in newspaper comment sections. Trolls have been known to create a hostile environment in comment sections, often motivated by attention seeking and amusement. In recent years, following the Brexit vote and the U.S. presidential election of 2016, trolls have also been accused of actively undermining the Western political climate by using social media to divide political opponents. Furthermore, technological development has led to the possibility of automated software, known as bots, playing a role in online debates. As social media users and participants of online comment sections become more digitally literate, the awareness of trolls and bots will hopefully make people less susceptible to online manipulation. But this awareness could also cause commenters to discredit and delegitimize opposing arguments in comment sections by accusing others of being a troll or a bot, without considering the merits of the argument itself. If this is the case, it constitutes a challenge in creating a democratically valuable debate in comment sections. In this study, comments from three U.S. news sites were sampled and analyzed to investigate how accusations of trolling are made, and how debates are affected by such accusations. The results showed that right-wing commenters were more likely to be accused of trolling, and that these accusations seem to have been motivated by political differences. Accusers would either challenge the suspected troll, critique the effectiveness of the perceived trolling, make fun of the suspected troll, or simply warn other commenters about their presence. Finally, while debates often continued after an accusation of trolling had been made, the accuser and the accused rarely participated further. The results suggest that accusations of trolling do not have any major impact on the debate. It is, however, problematic that such accusations seem to be used as a rhetorical tool to discredit opposing arguments, which could lower the deliberative quality of debates in comment sections.

Contents

Introduction

Newspaper comment sections have been described as a staple of the online experience (Finley, 2015). With approximately 90 percent of news sites having some form of comment section (Stroud, *et al.*, 2017) it has become possible for readers of almost any newspaper to share their views to a large audience and add their voices to public debates (Artime, 2016). Newspaper comment sections provide an arena for public debate and have been found to shape the opinions of the readers and influence how journalists work (Toepfl and Piwoni, 2015). As with any digital platform with user-generated content, newspaper comment sections can be susceptible to trolling behavior. In recent years, mainstream media has given trolls and bots much attention, including how trolling may be used as a method for political influencing. As Internet users become more knowledgeable about these disruptive elements, they may expect to encounter trolls in comment sections. The availability heuristic, a psychological mechanism in which a person judges the likelihood of an event by how readily pertinent examples come to mind [1], could possibly affect the likelihood of a commenter judging the author of a disagreeable comment as a troll. Overreporting of a topic can lead to individuals experiencing a biased assessment of risk [2]. Because of the increased mainstream reporting on trolling, bots and social media being used for foreign political influence, individuals may form a biased assessment of the risk of encountering trolls or bots online, including in newspaper comment sections.

The increased focus on trolling could cause users of comment sections to react appropriately to divisive content and trolling behavior. However, it may also provide an opportunity for debaters to disregard arguments from people with opposing political views. Accusations of trolling could potentially be used to shut down opposing arguments, whether these are made by trolls or not. In some cases, a commenter may even be accused of being a bot. This qualitative study aims to explore accusations of trolling in the comment sections of three newspapers: *Politico, Washington Post* and *New York Times*. Comment sections have the potential for being a democratically valuable forum for public debate, where individuals can openly discuss topics of common interest and share experiences and information relevant to news stories. Therefore, it’s important not only to understand how debates in comment sections are affected by trolling, but also how participants react to the possibility of trolling taking place. At its core, an accusation of trolling represents a disbelief in a commenter’s intentions and credibility, and it is important to understand the motivations and effects of such accusations. To explore this topic, this study will investigate how accusations of trolling in newspaper comment sections are made, and how these accusations are responded to.

Research on trolling in newspaper comment sections has several methodological challenges. Firstly, comment sections are usually moderated by newspaper employees who may delete comments containing examples of trolling. In recent years, newspapers have begun taking editorial action against unwanted comments, such as increased moderation, and identifying commenters by requiring them to sign up for an account or having them sign their comments using their Facebook identity (Gonçalves, 2015; Ihlebæk, *et al.*, 2013; Sonderman, 2011; Stroud, *et al.*, 2017).

Secondly, identifying comments that are written by trolls can be problematic. The term trolling can refer to a variety of online activities, some of which may look innocent at first glance. For example, trolls can share positive content to gain an online following (Linvill and Warren, 2019), or pretend to agree with the opposing side of an issue to voice their disagreements with that side in the form of "concerns" (Castile, 2016). Internet trolls can also be considered a form of social hackers who, according to Kerr and Lee (2019), uses technical and soft skills, such as manipulating social interactions and dynamics, to manipulate their targets. For most people, however, the term trolling usually refers to uncivil or impolite online behavior. But such behavior in comment sections could be confused with sincere but uncivil or impolite comments, which is commonly found in newspaper comment sections (Graham and Wright, 2015; Reagle, 2015; Rowe, 2015). While identifying comments written by trolls can be difficult, identifying accusations of trolling is less challenging. The current study investigates such accusations, to better understand how accusations of trolling affect the debate in comment sections. The study has three goals: 1) to analyze comments that have been accused of being written by trolls or bots, 2) to analyze how such accusations are made, and 3) to investigate how such accusations are responded to. In this paper, I will go through current research on the topics of trolling and bots, and the mechanisms by which the increased mainstream attention to these topics could make commenters more likely to judge opposing arguments in comment sections as trolling behavior. I will then explain the methodology and results of the study, before discussing the results.

Trolls and bots

The online world provides us with an unprecedented amount of information. However, as Hardaker points out, that information can be dangerously wrong, and computer-mediated communication involves the possibility of deception [3]. Deception is at the core of trolling, which has been defined as "the practice of behaving in a deceptive, destructive, or disruptive manner in a social setting on the Internet with no apparent instrumental purpose" (Buckels, *et al.*, 2014). Traditionally, trolls are jokesters who behave in an antagonistic way for their own amusement's sake [4]. Their attempts to elicit reactions from their victims can be motivated by boredom, attention seeking, revenge, pleasure, and a desire to cause damage to a community (Shachaf and Hara, 2010). There has also been found a correlation between trolling behavior and certain personality traits. Using a variety of personality tests such as the Short Sadistic Impulse Scale, Varieties of Sadistic Tendencies Scale, Comprehensive Assessment of Sadistic Tendencies, Short Dark Triad Scale, and Big Five Inventory, researchers found that trolling correlates with sadism, psychopathy, and Machiavellianism (Buckels, *et al.*, 2014). In a more recent study, Buckels, *et al.* (2019) found that trolls and sadists found pleasure in visual representations of people in physical or emotional pain, while downplaying the magnitude of that pain, and that trolls and sadists reacted more positively to reading about harmful scenarios.

In addition to trolling, bots have become a well-known online phenomenon. The term is defined by Bastos and Mercea as "automatic posting protocols used to relay content in a programmatic fashion" [5]. Bots are essentially computer programs that can use the Internet to add content to social media platforms. They have a wide variety of usage, including user interaction and automation of tedious tasks (Lebeuf, *et al.*, 2018). Bots created for interaction with humans have been found to lack authenticity and social competence (Neururer, *et al.*, 2018). It has been found, however, that bots can be used successfully to spread disinformation on Twitter. One study found that bots were used to spread anti vaccine messages on the social media platform (Broniatowski, *et al.*, 2018). The researchers found that the strategy used by trolls was to generate several tweets about the same topic to flood the discourse, and that the bots posted content at a higher rate than the average Twitter user. The bots were primarily used for spreading content, while the human trolls promoted discord by targeting both sides of the vaccine debate. Another study on the U.K. Brexit referendum found that bots on Twitter were effective at creating small- to medium-sized retweet cascades, that content retweeted by bots compromised user-generated hyperpartisan news, and that clusters of bots in a botnet could replicate active users (Bastos and Mercea, 2019).

A much-discussed topic in recent years is the idea of foreign influence on Western politics. Organized cyber operations have been used to influence European politics, though the effect of such activities is described as limited (Karlsen, 2019). According to Stewart, *et al.* (2018), troll accounts on Twitter took advantage of the Black Lives Matter movement to create discord during the U.S. presidential election of 2016. The content produced by these accounts rarely crossed political divides, suggesting that filter bubbles and echo chambers keep disinformation within political camps. These types of findings help to fuel a general conception of divisive political influence through social media, sometimes perpetrated by foreign entities.

Accusations of trolling

While the history of trolling can be traced back to the 1980s, the concept didn't receive much mainstream attention until 2010 [6]. In recent years, the topic of foreign influence on Western politics has received much attention by the mainstream media. Most people are aware of the existence of bots, if only because most Internet users will at one point have to prove their humanity by completing a captcha test to prove they're not a robot — a test that bots have been known to pass (Sulleyman, 2017). In addition, bots have received much media attention in later years, with one study about bots' influence on the 2016 Brexit referendum in the U.K. being reported on in over 250 news articles [7]. Terms such as trolling and bots have become widely used in digital communities, such as comment sections, and there is much awareness of these disruptive elements among internet users as anxiety about trust, facts, and democracy is intensifying (Dimock, 2019).

Having knowledge and understanding of online phenomena is considered by researchers as an important skill and a requirement for democratic participation, as well as by public institutions such as the European Union (European Commission, 2016). In the early days of the Web, Wang (1996) argued that educating the ignorant would help against the negative effects of flaming. Howard Rheingold writes that "those who understand the fundamentals of digital participation, online collaboration, informational credibility testing, and network awareness will be able to exert more control over their own fates than those who lack this lore." [8]. Graham and Wright (2015) are among the researchers who expect that user behavior have evolved as people gain more experience with, for example, trolling. Kerr and Lee (2019) claims that lack of technical literacy is one of the aspects of their targets that trolls take advantage of, meaning that increased technical literacy should make Internet users less susceptible to trolling.

When Internet users have more knowledge about trolls and bots, accusations of trolling are expected to increase. Accusations of trolling may function as a tool to delegitimize extremist point of views or actual trolling behavior in comment sections. However, they may also be used simply to discredit and delegitimize arguments one does not agree with. Having knowledge about disruptive elements such as trolls and bots may provide an opportunity for debaters to disregard opposing arguments by claiming they are made by people or bots with sinister intentions. In any online discussion, there will be disagreements. When faced with arguments that go against their preconceptions, a person may rationalize their beliefs by discrediting opposing arguments [9]. When having knowledge about the existence of trolls and bots, a person can discredit and delegitimize an opposing argument by accusing its author of being a troll or a bot, without having to consider the merits of the argument itself. If, for example, a person who identifies as a liberal sees a comment that they find offensive because it's written in support of conservative ideals, that person may be tempted to think the comment is written by a troll simply because they are aware of the issues with trolling from mainstream media. This may be problematic in creating a democratically valuable online debate, as accusations of trolling could become a form of exclusion that decreases the value of online political debates. It may also be problematic on a personal level for any real person making an argument, only to be met with accusations of being a troll or a foreign agent, or not even being human. It could be uncomfortable for a person to have their arguments dismissed, and to be accused of being something that they are not.

There has been little research on how accusations of trolling are responded to by the person being accused, or by other commenters. This has caused a gap in our understanding of online debates. Comment sections are the target of much research on how incivility and toxic disinhibition affects their deliberative value. But I would argue that if accusations of trolling are used as a rhetorical tool to devalue opposing arguments, this could also affect the deliberative value of comment sections. However, despite the lack of research into accusations of trolling, some research has been done on how trolls are responded to by others. Hardaker considered accusations of trolling on Usenet and identified seven types of responses to trolling behavior: 1) Engaging by responding sincerely to the troll; 2) Ignoring the trolling attempt; 3) Exposing the troller to the rest of the group; 4) Challenging the troller directly or indirectly; 5) Critiquing the effectiveness, success, or quality of the troller; 6) Mocking or parodying the trolling attempt; and, 7) Reciprocating by trolling the troller [10]. Hardaker's study focuses on creating a taxonomy of different ways people respond to perceived trolling, which makes it interesting in the current study. Hardaker's response types is one of the methods that will be used in the current study to investigate accusations of trolling in comment sections.

Methodology

The three newspapers chosen for this study were *Politico, Washington Post*, and *New York Times*. These newspapers were chosen because they provide different venues for studying comment sections, with different levels of anonymity. *Politico* is a free-to-read newspaper that uses a Facebook plug-in as a comment section. This means that commenters on *Politico* must use their Facebook account when commenting. The *Washington Post* and *New York Times* do not use Facebook for their comment sections. Commenters on these news sites must create an account and choose a username, which can either be a pseudonym or their real name. The *Washington Post* and *New York Times* also have online subscription models that pose a barrier for some commenters, as a subscription is required to be able to read and comment on any article.

Constructed week sampling was used to create two constructed weeks from February of 2018 to February of 2019 for each newspaper being studied. This involved selecting two random Mondays, two random Tuesdays, etc., during the specified timeframe. This method of sampling is recommended for studying daily newspapers because it creates a randomly selected issue for each day of the week. Two constructed weeks have been found to be sufficient for representing a year's content [11]. A total of 3,851 comments were collected from politically themed articles and stored in a database using this method. To ensure the anonymity of the commenters, names were continuously replaced with numeric identifiers.

After the comments were sampled, search queries were devised to identify accusations of trolling. Through a combination of SQL-queries and free search, comments containing any combination of the words "troll", "bot", and "Russian" were identified. While this was a thorough method for searching the comments, it does not guarantee that all accusations were found. Some misspelled words or accusations using unknown analogies may have been missed.

To analyze the sampled comments, a descriptive approach was used for each identified case (*n*=24). First, the different commenters and their roles were established; accused commenter, accuser, and other commenters. Then the accused and accuser's comments were analyzed carefully for further details about how they communicate, and the seven response types identified by Hardaker (2015) were used to categorize the accusations of trolling. While these response types are not a crucial part of the current study, I would argue that the incorporation of existing taxonomies could serve a function by highlighting aspects of the data that I would not have considered otherwise. Hardaker's taxonomy was used because it provides established categories for identifying different types of responses to perceived trolling. Finally, the general discussion was mapped out with special emphasis being put on how the different commenters respond to accusations of trolling and how such accusations affected the discussion. After having described each case, general trends were identified.

This methodology was approved by the Norwegian Centre for Research Data (*Norsk senter for forskningsdata*), which has imposed constraints to protect the privacy of the commenters whose data has been sampled. Even anonymized datasets can contain personal information that can cause a person to be identifiable (Markham and Buchanan, 2012). Therefore, in the following presentation of the results of this study, no comments will be quoted. Paraphrasing and descriptions will instead be used to illustrate the findings.

Results

In total, 30 accusations of trolling were found in the studied data, written by 31 accusers, and directed at 24 accused commenters. The reason for the discrepancy between the number of accusations and accusers is that one commenter accused someone of being both a troll and a bot. 24 (1.71 percent) of the comments from *Politico* contained some form of accusation, while only five (0.35%) from the *Washington Post* and one (0.09%) from the *New York Times* contained accusations. In other words, *Politico* had far more accusations of trolling in its comment sections than the other two news sites. This could be because *Politico* is the only Web site of the three that uses a Facebook plug-in for their comment sections. However, the observed difference could also be explained by demographic differences between the commenters on the different news sites. It is also worth noting that both the *Washington Post* and the *New York Times* have several barriers for commenting that *Politico* does not. Both papers require users to create a dedicated account on their Web sites to be able to comment. In addition, they have subscription plans that limit the activity of non-paying readers. This may create a barrier for trolls, which inadvertently reduces the number of accusations of trolling.

The accusations of trolling showed great variation in length, argumentative and rhetorical style, as well as temperament. Some of them were short — sometimes one-word long accusations of someone being a troll or a bot. Others were longer and argumentative. At times, an accuser seemed agitated by the perceived trolling, while other accusers seemed to find amusement in it. Some comments were directed at the person being accused of trolling, while some were directed at other commenters.

Table 1: Overview of the results, where each row represents a case of someone being accused of trolling or of being a bot.
Note: *Political leaning in this context refers to whether the accused commenter had expressed views in favor of a political side in American politics and may not reflect the right-to-left political spectrum of other countries and regions.

Case	Number of accusations	Political leaning of accused*	Continued discussion after accusation	Accused replies or continues to participate in discussion	Words used about the accused commenter	Type of response (Hardaker)
1	1	Left	Yes	No	Troll/Bot	3
2	1	Left	Yes	No	Bot	3
3	1	Right	Yes	Yes	Russian troll	3

4	1	Left	Yes	No	Bot	3
5	1	Left	Yes	No	Troll	6
6	1	Right	No	No	Troll	5
7	1	Right	Yes	No	Troll	3
8	1	Right	Yes	Yes	Russian troll	4
9	1	Right	Yes	No	Russian troll	3
10	1	Right	Yes	Yes	Troll	3
11	2	Left	Yes	No	Russian troll/Troll	6,4
12	1	Right	Yes	Yes	Troll	4
13	1	Right	Yes	No	Troll	5
14	1	Right	Yes	No	Russian Troll	3
15	5	Right	Yes	Yes	3x Troll/2x Bot	5,5,3,3,3
16	1	Right	Yes	No	Troll	6
17	1	Right	Yes	No	Troll	4
18	1	Right	Yes	Yes	Troll	3
19	1	Right	No	No	Russian Troll	5
20	2	Right	Yes	No	Russian Troll/Troll	3,6
21	1	Right	No	No	Troll	3
22	1	Right	No	No	Troll	3
23	1	Right	No	No	Russian Troll	5
24	1	Right	No	No	Troll	4
Total	30	R: 19, L: 5	Yes: 18, No: 6	Yes: 6, No: 18	T: 18, RT: 6, B: 5	

Trolls, Russian trolls, and bots

As can be seen in Table 1, 18 of the accusations made were accusations of trolling. In addition to this, there were eight accusations of someone being specifically a Russian troll, and five accusations of someone being a bot. As illustrated by the word cloud in Figure 1, the most common accusation was when a commenter used the word troll. Sometimes this was the only word found in a comment, but mostly it was used within a sentence. It was typical for the accuser to address the accusation to other commenters. A typical example is when an accuser writes that other commenters should “ignore this troll”, or “he is obviously a troll”. Other times, the accuser directed the comment at the accused. Examples of such accusations are “Get lost, troll” or “Do you really believe that people believe your crap, troll?” On two occasions, the word troll was used in combination with a hashtag: #faketrumptroll and #purgethetrolls.

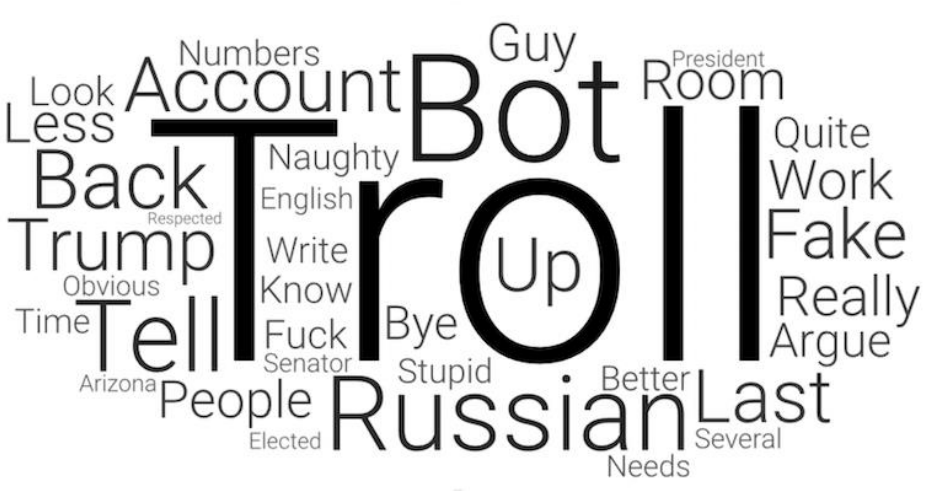


Figure 1: Word cloud of the most used words in comments accusing someone of trolling.

Accusations of someone being a Russian troll followed a similar pattern as described above, but with some added rhetoric. These accusations were sometimes used in combination with other derogatory rhetoric. One commenter was accused of being a “*Russian teenage troll*” that ate too many potato chips. Another was called a sock puppet of the Russian government, and yet another was called a whore in addition to being accused of being a Russian troll.

Accusations of someone being a bot were made in a slightly different way. Mainly, none of these accusations were directed at the accused commenters. Instead, they were seemingly written to inform other commenters that the accused was believed to be a bot, by writing something along the lines of “*Do not engage with this commenter because it’s a bot*”. Following the logic of the person making the accusation, this makes sense. There would be no point in calling out a bot by engaging with it as if it was a real human being.

Types of accusations

The types of accusations were investigated using Hardaker's (2015) response types. Most replies made to commenters who were accused of trolling would be categorized as sincere engagement, as most commenters would argue with the accused troll without making any accusations. However, most of the accusations of trolling would fall into one of four of Hardaker's response types to trolling: Exposing the troll to the rest of the group, challenging the troller, critiquing the troll, and mocking or parodying the troll attempt (Table 2).

Table 2: Results of coding using Hardaker's (2015) types of responses to perceived trolling.

(3) Exposing the troller to the rest of the group	15
(4) Challenging the troller directly or indirectly	5
(5) Critiquing the effectiveness, success, or quality of the troller	6
(6) Mocking or parodying the trolling attempt	4

As can be seen in Table 2, half of the accusations of trolling fell within category 3 of Hardaker's types of responses to perceived trolling. It seems the most common way to accuse someone of trolling is by informing the rest of the commenters about the perceived troll. In one case, the accuser wrote that the accused is "*probably*" just a bot or a troll, and then went on to explain how one can tell if a commenter is not being sincere. The accuser argued that a troll's goal is to cause division and anger in the comment sections. In another case, the accuser wrote that the accused was using a fake Facebook account, arguing that this indicated that they must have been a Russian troll.

Another observed type of response to perceived trolling was when accusers challenged the accused commenter. In one case, the accuser told a commenter that he identified as a Russian troll to "*get lost*". In another case, the accused was called a "*coward*" and a "*zero*" because he was trolling. There were also several instances of the accuser saying that the accused commenter "*sounds like a troll*".

Several accusers would use critique to call out perceived trolling. One such case involved the accuser calling a commenter a Russian troll because he had misunderstood the difference between American 800 and 900 phone numbers. Telephone numbers starting with the prefix 800 are usually toll-free in the U.S., while 900 numbers are identified as premium-rate telephone numbers, because additional services are provided (U.S. Federal Communications Commission, 2019). The accuser pointed out that any real American should know the difference between the two. This was sufficient for the accuser to believe that the accused was not a real American. Another such case was when an accuser thought that the accused commenter's English skills were not good enough for him or her to be a native English speaker, and therefore the commenter must have been a Russian troll. The remaining accusations in this category simply contained phrases such as "*low-effort trolling*" or "*you trolls are becoming pathetic*".

As mocking can be used for criticism, the final observed type of responses, *mocking or parodying the trolling attempt*, is similar to the previous category. But in these cases, the accuser seemed to have found some humorous enjoyment in the perceived attempts at trolling. One such commenter wrote "*hahaha*" before asking if the troll expected people to believe him. In another case, the accuser told the accused that he found the trolling hilarious, before asking the commenter to go back to troll school.

Vulgar, divisive, or conspiratorial comments

Of the 24 comments that were met with accusations of trolling, about half (*n*=13) were divisive, conspiratorial, or contained vulgarity. These were comments written by commenters who showed an attitude of non-cooperation and divisiveness. The remaining comments were more argumentative or informative but had a clear political leaning.

Vulgar comments were easily identified because they contained some form of vulgar language, usually derogatory curse words directed at other commenters or political figures. Divisive comments were those that displayed hostility and non-cooperativeness. An example of this was when a commenter accused of trolling wrote a very hostile comment about Canada. In another case, the accused commenter wrote that Trump supporters were psychopaths.

Conspiratorial comments were written by commenters sharing conspiracy theories, such as when one commenter accused of trolling wrote about collusion between the FBI and the Democratic Party. Another commenter accused of trolling wrote that President Obama had given Iran nuclear weapons, and even specified an exact number of weapons that had been given.

Political differences

Commenters accusing others of trolling rarely explained why they made accusations. Five of the accusers in this study specifically explained that the accused person's Facebook account seemed fake. In addition, as noted above, two accusations were based on poor English skills or not understanding a particular part of American culture. Most accusations, however, contained no such explanation. In combination with the fact that half of the accused commenters did not write obviously divisive or vulgar comments, it seems that many accusations of trolling were made because of political disagreements. Most accusations of trolling were made towards commenters expressing politically right-wing views, *i.e.*, commenters who showed support for Republicans and/or criticized Democrats. Twenty-five accusations were made by left-wing commenters, directed towards 19 right-wing commenters. In comparison, only five accusations of trolling were made by right-wing commenters.

How accusations of trolling affect the debate

Accusations of trolling were rarely replied to by the accused person or other commenters. The general trend seems to be that such accusations are ignored by other commenters, as the discussion tends to continue as before without any further comments being made by the accused or the accuser. This suggests that accusations of trolling do not have much effect on the discussion. Only seven times did a person being accused of trolling continue the discussion. Of these, only one addressed the actual accusation of trolling. This commenter, who had been accused of being a Russian troll, tagged the accuser in a response where he wrote "*Nice try*".

Discussion

The current study had three goals: to analyze comments that were accused of being written by trolls or bots, to analyze how such accusations are made, and to investigate how these accusations are responded to. While the scope of this study is limited, it does show a pattern that allows for the following conclusions to be made about the data:

1. *Most accusations of trolling are made by left-wing commenters and directed towards commenters expressing right-wing views*. This suggests that there is a political divide between those accusing and those being accused of trolling. There may be several explanations for this finding. Firstly, it may be that right-wing commenters more often behave in a way that would elicit accusations of trolling. However, as I will discuss in more detail, this may not be true in all cases because only half of the accused commenters wrote divisive or vulgar comments. Secondly, it may be that the mainstream attention on trolling influence commenters' expectations about who could be a troll. The possibility of foreign influence on the 2016 U.S. presidential election has been given much mainstream attention, and much of this coverage has focused on trolling and fake news possibly helping the right-wing candidate to win. This could make left-wing commenters more suspicious about the intentions of right-wing commenters and make them more likely to accuse right-wing commenters of trolling.
2. *It is common for accusations of trolling to be motivated by political differences*. Half of the accused commenters wrote comments that were clearly divisive, conspiratorial, or vulgar. While this may seem like trolling behavior, it is worth noting that many similar comments were written by commenters who were not accused of trolling. The other half of the accused commenters wrote argumentative or informative comments expressing their opinions in a way that might be expected in a comment section. In addition, many accusers also argued against the views of the accused commenters, and only a few of them specifically made claims of fake profiles when explaining the reason for their accusations. It seems therefore that most accusations of trolling were made because of a political disagreement. This finding is particularly troubling, because it suggests that people with a certain political viewpoint are at higher risk of having their arguments dismissed with accusations of trolling. When arguments in favor of a

right-wing opinion are dismissed without being challenged by opposing argumentation, the deliberative value of a given debate suffers.

3. *Most of the commenters accusing someone of trolling will either challenge the accused troll's arguments, mock or critique the troll, or warn other commenters about the presence of a troll*. This conclusion was made using Hardaker's response types to perceived trolling [12]. Four of her seven categories were identified in the current study; 1) Exposing the troller to the rest of the group; 2) Challenging the troller directly or indirectly; 3) Critiquing the effectiveness, success, or quality of the troller; and, 4) Mocking or parodying the trolling attempt. It is difficult to say why only four out of seven response types were identified in this study. It could be that there was a lack of data for the remaining three categories to be identified, that there were differences between the coders, differences between the platforms being studied, or that Hardaker created too many and too narrow categories. It should also be noted that some of Hardaker's response types can overlap or blend together. An example of this is the response type *Critiquing the effectiveness, success, or quality of the troll*. Such critique can often be expressed by *mocking or parodying the trolling attempt* — which is a different response type.
4. *Accusations of trolling are rarely responded to by the accused person or other commenters*. In only a few cases did the accused person continue the discussion, and in only one of them did the accused person confront the accusation of trolling. It would be tempting to suggest that the lack of further commenting from people accused of trolling suggests that the accusation has discouraged them from further participation. However, there are several other possibilities; perhaps they never intended to write more than the one comment, perhaps they never even saw the accusation, or perhaps they were indeed trolls. What is certain, however, is that other commenters mostly ignored accusations of trolling. Even when the accuser specifically encourages them to ignore the perceived troll, the discussion tends to continue as before. This suggests that accusing someone of being a troll does not discourage other commenters from engaging with the troll, and that such accusations have little effect on the debate. This would mean that false accusations of trolling are mostly ignored, but also that any legitimate accusations will not discourage others from engaging with the troll.
5. *Accusations of trolling were more common on Politico*. *Politico* had by far the most accusations of trolling in their comment section. *Politico* is the only one of the three news sites that use a Facebook plug-in as a comment section. This means that anyone with a Facebook account can easily comment without having to create a separate account, as opposed to the *Washington Post* and the *New York Times* who use their own comment section plugins that require the creation of a separate account. While this is only speculation, it could be that the lower barrier for commenting on *Politico* leads to more actual trolling and more accusations of trolling. Furthermore, fake Facebook accounts have been discussed in mainstream media (Kottasová, 2017; Shane and Goel, 2017; Weise, 2017), which could lead to distrust in commenters using Facebook for identification. It is worth noting, however, that this study has only investigated three news sites and cannot make

definitive conclusions about how the type of comment section being used affects the frequency of accusations of trolling.

Limitations of the study

The results of this study do not reveal a complete picture of the topic of trolling accusations in newspaper comment sections. The relatively few cases do not allow for broad conclusions to be made. A more quantitative study, using content analysis to categorize and quantify different types of accusations and responses, could shed further light on this topic. Another problem with the current study is that it has been difficult to validate the identities of people accused of having fake Facebook accounts. This is because of the technical and ethical limitations of the study that required the data to be anonymized in such a way that further investigations into the commenters themselves was impossible. Finally, it is worth mentioning that when studying data from comment sections, the data may be incomplete. Some comments that could have shed more light on the subject may have been deleted, ether by the commenters themselves or by moderators.

Conclusion

This study has explored a topic of research that has received little previous attention. As discussed in the literature review, while trolling has been a topic of research, little attention has been given to accusations of trolling. I have theorized that the increased mainstream attention to topics like trolling, foreign political influence, and bots can lead to individuals becoming more aware of these concepts. This in turn could lead to them identifying certain behaviors as trolling, whether or not they are caused by actual trolls. Newspaper comment sections, where strangers engage in political debates, is an arena where this can happen. Political disagreements may lead to accusations of trolling, and such accusations could be used as a rhetorical tool to dismiss opposing arguments. If this were true, it would constitute a challenge to the deliberative value of comment sections. By using real-world examples of comments from *Politico, Washington Post*, and *New York Times*, this study has analyzed accusations of trolling and how such accusations affect the debate.

This study has uncovered trends that have led to several conclusions about how accusations of trolling affect the debate in newspaper comment sections. Accusations of trolling often targeted right-wing commenters, were made because of political disagreements, were rarely responded to by the accused, and were mostly ignored by other commenters as the debates continued. If these conclusions were confirmed by further research, they would further illuminate a topic of public interest; the democratic value of comment sections. If comment sections are to serve as a forum for public debate, inclusion and openness should be valued. The activities of trolls, real or imaginary, and how they are responded to, can affect how people communicate in comment sections, the trust between commenters, and the inclusion of all those who want to participate. FM

About the author

Magnus Knustad is a Ph.D. candidate in digital culture at the University of Bergen (*Universitetet i Bergen*), researching comment sections on news articles.
E-mail: magnus [dot] knustad [at] uib [dot] no

Notes

[1] Gilovich, *et al.*, 2016, p. 137.

[2] Gilovich, *et al.*, 2016, p. 139.

[3] Hardaker, 2010, p. 223.

[4] Hardaker, 2015, p. 202.

[5] Bastos and Mercea, 2018, p. 2.

[6] Hardaker, 2015, p. 202.

[7] Bastos and Mercea, 2018, p. 2.

[8] Rheingold, 2012, p. 2.

[9] Gilovich, *et al.*, 2016, p. 239.

[10] Hardaker, 2015, p. 223.

[11] Riffe, *et al.*, 2014, pp. 85&nndash;86.

[12] Hardaker, 2015, p. 223.

References

M. Artime, 2016. “Angry and alone: Demographic characteristics of those who post to online comment sections,” *Social Sciences*, volume 5, number 4.
doi: https://doi.org/10.3390/socsci5040068, accessed 17 July 2020.

M.T. Bastos and D. Mercea, 2019. “The Brexit botnet and user-generated hyperpartisan news,” *Social Science Computer Review*, volume 37, number 1, pp. 38–54.
doi: https://doi.org/10.1177/0894439317734157, accessed 17 July 2020.

M. Bastos and D. Mercea, 2018. “The public accountability of social platforms: Lessons from a study on bots and trolls in the Brexit campaign,” *Philosophical Transactions of the Royal Society A*, volume 376, number 2128 (13 September).
doi: https://doi.org/10.1098/rsta.2018.0003, accessed 17 July 2020.

D.A. Broniatowski, A.M. Jamison, S. Qi, L. Alkulaib, T. Chen, A. Benton, S.C. Quinn, and M. Dredze, 2018. “Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate,” *American Journal of Public Health*, volume 108, number 10, pp. 1,378–1,384.
doi: https://doi.org/10.2105/AJPH.2018.304567, accessed 17 July 2020.

E.E. Buckels, P.D. Trapnell, and D.L. Paulhus, 2014. “Trolls just want to have fun,” *Personality and Individual Differences*, volume 67, pp. 97–102.
doi: https://doi.org/10.1016/j.paid.2014.01.016, accessed 17 July 2020.

E.E. Buckels, P.D. Trapnell, T. Andjelovic, and D.L. Paulhus, 2019. “Internet trolling and everyday sadism: Parallel effects on pain perception and moral judgment,” *Journal of Personality*, volume 87, number 2, pp. 328–340.
doi: https://doi.org/10.1111/jopy.12393, accessed 17 July 2020.

E. Castile, 2016. “Watch out for this kind of troll,” *Bustle* (26 February), at https://www.bustle.com/articles/144447-what-is-concern-trolling-watch-out-for-this-subtle-form-of-shaming, accessed 17 July 2020.

M. Dimock, 2019. “An update on our research into trust, facts and democracy,” *Pew Research Center* (5 June), at https://www.pewresearch.org/2019/06/05/an-update-on-our-research-into-trust-facts-and-democracy/, accessed 17 July 2020.

European Commission, 2016. “Digital Skills at the core of the new Skills Agenda for Europe” (10 June), at https://ec.europa.eu/digital-single-market/en/news/digital-skills-core-new-skills-agenda-europe, accessed 17 July 2020.

K. Finley, 2015. “A brief history of the end of the comments,” *Wired* (8 October), at https://www.wired.com/2015/10/brief-history-of-the-demise-of-the-comments-timeline/, accessed 17 July 2020.

T. Gilovich, D. Keltner, S. Chen, and R.E. Nisbett, 2016. *Social psychology*. Fourth edition. New York: W.W. Norton.

J. Gonçalves, 2015. “A peaceful pyramid? Hierarchy and anonymity in newspaper comment sections,” *Observatorio*, volume 9, number 4, pp. 1–13, and at at http://www.scielo.mec.pt/scielo.php?lng=en, accessed 17 July 2020.

T. Graham and S. Wright, 2015. “A tale of two stories from ‘Below the Line’: Comment fields at the *Guardian*,” *International Journal of Press/Politics*, volume 20, number 3, pp. 317–338.
doi: https://doi.org/10.1177/1940161215581926, accessed 17 July 2020.

C. Hardaker, 2015. "'I refuse to respond to this obvious troll": An overview of responses to (perceived) trolling," *Corpora*, volume 10, number 2, pp. 201–229.
doi: https://doi.org/10.3366/cor.2015.0074, accessed 17 July 2020.

C. Hardaker, 2010. "Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions," *Journal of Politeness Research*, volume 6, number 2.
doi: https://doi.org/10.1515/jplr.2010.011, accessed 17 July 2020.

K.A. Ihlebæk, A.S. Løvlie, and H. Mainsah, 2013. "Mer åpenhet, mer kontroll: Håndteringen av nettdebatten etter 22," *Norsk Medietidsskrift*, volume 20, number 3, pp. 223–240, and at https://www.idunn.no/nmt/2013/03/mer_aapenhet_mer_kontroll_-_haandteringen_av_nettdebatten_e, accessed 17 July 2020.

G.H. Karlsen, 2019. "Divide and rule: Ten lessons about Russian political influence activities in Europe," *Palgrave Communications*, volume 5, article number 19.
doi: https://doi.org/10.1057/s41599-019-0227-8, accessed 17 July 2020.

E. Kerr and C.A.L. Lee, 2019. "Trolls maintained: Baiting technological infrastructures of informational justice," Information, Communication & Society (28 May).
doi: https://doi.org/10.1080/1369118X.2019.1623903, accessed 17 July 2020.

I. Kottasová, 2017. "Facebook targets 30,000 fake accounts in France," *CNN* (21 April), at https://money.cnn.com/2017/04/14/media/facebook-fake-news-france-election/, accessed 17 July 2020.

C. Lebeuf, M.-A. Storey, and A. Zagalsky, 2018. "Software bots," *IEEE Software*, volume 35, number 1, pp. 18–23.
doi: https://doi.org/10.1109/MS.2017.4541027, accessed 17 July 2020.

D. Linvill and P. Warren, 2019. "That uplifting tweet you just shared? A Russian troll sent it," *Rolling Stone* (25 November), at https://www.rollingstone.com/politics/politics-features/russia-troll-2020-election-interference-twitter-916482/, accessed 17 July 2020.

A. Markham and E. Buchanan, 2012. "Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee," version 2.0, at https://aoir.org/reports/ethics2.pdf, accessed 17 July 2020.

M. Neururer, S. Schlögl, L. Brinkschulte, and A. Groth, 2018. "Perceptions on authenticity in chat bots," *Multimodal Technologies and Interaction*, volume 2, number 3.
doi: https://doi.org/10.3390/mti2030060, accessed 17 July 2020.

J.M. Reagle, 2015. *Reading the comments: Likers, haters, and manipulators at the bottom of the Web*. Cambridge, Mass.: MIT Press.

H. Rheingold, 2012. *Net smart: How to thrive online*. Cambridge, Mass.: MIT Press.

D. Riffe, S. Lacy, and F. Fico, 2014. *Analyzing media messages: Using quantitative content analysis in research*. Third edition. New York: Routledge.

I. Rowe, 2015. "Civility 2.0: A comparative analysis of incivility in online political discussion," *Information, Communication & Society*, volume 18, number 2, pp. 121–138. doi: https://doi.org/10.1080/1369118X.2014.940365, accessed 17 July 2020.

P. Shachaf and N. Hara, 2010. "Beyond vandalism: Wikipedia trolls," *Journal of Information Science*, volume 36, number 3, pp. 357–370. doi: https://doi.org/10.1177/0165551510365390, accessed 17 July 2020.

S. Shane and V. Goel, 2017. "Fake Russian Facebook accounts bought $100,000 in political ads," *New York Times* (6 September), at https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html, accessed 17 July 2020.

J. Sonderman, 2011. "News sites using Facebook Comments see higher quality discussion, more referrals," *Poynter* (18 August), at https://www.poynter.org/reporting-editing/2011/news-sites-using-facebook-comments-see-higher-quality-discussion-more-referrals/, accessed 17 July 2020.

L.G. Stewart, A. Arif,and K. Starbird, 2018. "Examining trolls and polarization with a retweet network," *Proceedings of WSDM Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*, at https://faculty.washington.edu/kstarbi/examining-trolls-polarization.pdf, accessed 17 July 2020.

N.J. Stroud, A. Muddiman, and J.M. Scacco, 2017. "Like, recommend, or respect? Altering political behavior in news comment sections," *ew Media & Society*, volume 19, number 11, pp. 1,727–1,743. doi: https://doi.org/10.1177/1461444816642420, accessed 17 July 2020.

A. Sulleyman, 2017. "Bot 'break' captcha, making the most annoying thing on the Internet pointless," *Independent* (31 October), at https://www.independent.co.uk/life-style/gadgets-and-tech/news/captcha-puzzles-recaptcha-solve-problems-vicarious-bots-artificial-intelligence-a8029401.html, accessed 17 July 2020.

F. Toepfl and E. Piwoni, 2015. "Public spheres in interaction: Comment sections of news Websites as counterpublic spheres," *Journal of Communication*, volume 65, number 3, pp. 465–488. doi: https://doi.org/10.1111/jcom.12156, accessed 17 July 2020.

U.S. Federal Communications Commission, 2019. "Pay-per-call information services" (31 December), at https://www.fcc.gov/consumers/guides/faqs-900-number-pay-call-services-and-fees, accessed 17 July 2020.

H. Wang, 1996. “Flaming: More than a necessary evil for academic mailing lists,” *Electronic Journal of Communication*, volume 6, number 1, at http://www.cios.org/EJCPUBLIC/006/1/00612.HTML, accessed 17 July 2020.

E. Weise, 2017. “Russian fake accounts showed posts to 126 million Facebook users,” *USA Today* (1 November), at https://eu.usatoday.com/story/tech/2017/10/30/russian-fake-accounts-showed-posts-126-million-facebook-users/815342001/, accessed 17 July 2020.

Editorial history

Received 9 September 2019; revised 7 June 2020; revised 9 June 2020; accepted 28 June 2020.

Get lost, troll: How accusations of trolling in newspaper comment sections affect the debate
by Magnus Knustad.
First Monday, Volume 25, Number 8 - 3 August 2020
https://journals.uic.edu/ojs/index.php/fm/article/download/10270/9576
doi: http://dx.doi.org/10.5210/fm.v25i8.10270

NORDIC ECOLABEL 241 PRINTED MATTER 731

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

UNIVERSITAS BERGENSIS

uib.no

ISBN: 9788230866443 (print)
9788230849217 (PDF)