The Antagonistic Battle between 'Good' and 'Evil'

A qualitative analysis of the interplay between digital hate culture and civil society counter efforts in the comment sections of Facebook

Nadine Keller

Abstract

Departing from the increasing threat that organized hate groups and their manipulative practices pose to contemporary society, this thesis seeks to unravel the workings of digital hate culture and to highlight the potential of civil society-led counter initiatives to combat the spread of hatred online. The research is based on a twofold qualitative content analysis. In a first step, the intended practices of two opposing groups – an organized hate group (*Reconquista Germanica*) and an organized counter speech group (*Reconquista Internet*) – are analyzed based on a set of internal strategic communication documents. In a second step, three comment threads on Facebook are examined to illustrate the actualized practices of users spreading hate and users who counter-speak. By drawing on a four-dimensional framework, the analysis thereby considers how practices, discourses, power relations, and the technological affordances of Facebook shape this interplay. With theoretical reference to Mouffe's (2005) work on the antagonistic nature of *the political* and today's post-political *Zeitgeist*, this thesis ultimately comes to discuss whether such confrontations between exponents of digital hate culture and counter speakers must be understood as irrefutable antagonisms or if productive agonism can be fostered through a mutual understanding of one another as legitimate adversaries.

What the analysis evinces is that the discussions carried out between the two opposing camps are highly moralized, resulting in an antagonistic battle between 'good' and 'evil' that interferes with the possibility for productive agonism. It is further shown that, in this post-political discussion climate, counter speech carries a crucial responsibility to conform to moral values and maintain professional and ethical standards to set itself apart from the harmful practices of digital hate culture. Otherwise, as the analysis indicates, counter efforts are likely to spur on destructive dynamics, further hardening the fronts between opposing positions that characterize today's increasingly polarized societies.

Keywords: agonism, antagonism, counter speech, digital hate culture, right-wing extremism, hate speech, Facebook, Mouffe, qualitative content analysis, user comments

Table of Contents

List of Figures

List of Tables

1 Introduction

Stereotyping and hostile portrayals of people based on different identity categories, such as race, religion, and gender have existed long before the digital era. Today, however, the Internet allows to spread narratives of hatred fast and globally, escalating the dangerous potential such rhetoric poses to targeted groups and society at large. Mainstream social media have become a primary tool for the organized dissemination of hate to large audiences (Blaya, 2018). In recent years, Europe and the US have witnessed an upsurge of right-wing ideology and a noticeable increase in openly displayed xenophobic, nationalist, racist, and anti-Semitic attitudes. Especially the hatred towards Muslims and Islam has become a prominent part of public discourse. When peeking into the comment sections of social media platforms, one can easily spot the abundance of hostile and derogative user comments, falsely suggesting that a majority of Internet users hold Islamophobic attitudes. Such distortion of public opinion is a major objective of organized groups constitutive of today's proliferating digital hate culture. Their denigrative and manipulative practices cannot only have far-reaching consequences for the emotional and physical well-being of targeted groups but also pose a threat to pluralist democracy. The culture of digital hate promotes discrimination and intolerance which accelerates the polarization of societies (Rieger, et al., 2018; George, 2015), making it a condition that requires urgent action. Although online hate speech and the spread of misinformation have been recognized as pressing issues among legal institutions and policy-makers in the last years (Awan, 2016), the current lack of efficient regulations makes countering by the individual user all the more important (Leonhard, et al., 2018). So far, however, we know little about the impact of such efforts (Blaya, 2018).

Departing from this knowledge gap, this thesis seeks to understand the different dimensions shaping the interplay between exponents of digital hate culture and *counter speakers*[1] on social media, striving to unravel the potential that civil society-led counter initiatives hold to work against the culture of digital hate. Anchoring this research in the workings of two opposing groups that are active in Germany, I first investigate the strategic communication practices of both organized hate speech and counter speech. On the one side stands *Reconquista Germanica* (RG), an organized far-right 'troll army' that operates under-cover to plot and execute cyberhate attacks against political opponents.

---

[1] The term is used to refer to those users who actively speak out against hateful content online.

On the other side stands *Reconquista Internet* (RI), a civil society movement that was established as a response to the disclosure of RG, aiming to "bring reality back to the Internet" and combat digital hate culture with "love and reason" (Reconquista Internet, 2018). With the help of a qualitative content analysis of the internal strategic communication documents of both groups, I unfold the strategies and tactics that underlie the practices of organized cyberhate and civil society counteraction. As the second step of the analysis, I examine three comment threads on Facebook belonging to journalistic articles that deal with the topics of immigration and Islam, as anti-refugee and anti-Muslim sentiments remain the most prominent nexus point across right-wing groups spreading hate online (Davey and Ebner, 2017). Conducting a qualitative content analysis, I investigate the practices of digital hate culture and counter efforts present in these user discussions. To ensure a holistic approach that goes beyond a mere textual analysis, I pay critical attention to the dimensions of discourse, power relations, and technological affordances and outline how different dynamics and factors affect the interplay of the two camps competing over the 'true' representation of the topics discussed.

With theoretical reference to the work of political scientist Chantal Mouffe (2005) on the antagonistic nature of *the political* and today's post-political *Zeitgeist*, the analysis examines whether such confrontations must be understood as irrefutable antagonisms that are harmful to society or if the discussions between the two opposing sides can be of productive agonistic nature and thus supportive and necessary for a healthy pluralist democracy. In this respect, it is crucial to not only assess the potential of counter efforts to be an effective means of citizen-led governance of digital hate culture but also to be sensitive to the risks counteraction carries to further pave the way for the spread of hatred.

With these objectives in mind, this thesis sets out to answer the following research questions:

How can we understand the interplay between digital hate culture and counter efforts on social media in terms of its potential to facilitate productive agonism?

➢ *What characterizes the communication strategies of organized digital hate culture and organized counter speech and how can they be understood within broader commenting practices on Facebook?*

> *What discourses and counter-discourses are fostered by hate speech and counter efforts?*
> *What role do the technological affordances of Facebook and power relations play in such confrontations?*

With the current upsurge of research on the rapid proliferation of organized hatred throughout the Internet and the relative lack of research on effective ways of countering it, this thesis steps into a timely and highly relevant discussion. With the insights acquired, this research ultimately seeks to contribute to an understanding of digital hate culture in online environments which can help undermine it and enhance counter initiatives. After providing a contextualization of the research topic (chapter 2) and a review of previous literature (chapter 3), the theoretical foundations are presented introducing Mouffe's theory on *the political* as well as the four-dimensional model that serves as an analytical framework (chapter 4). Thereafter, the qualitative methodological approach is mapped out (chapter 5) and ethical considerations are reflected on (chapter 6). With the presentation of the analysis (chapter 7), the thesis comes to discuss the proposed research questions before ending with a final discussion and concluding remarks (chapter 8).

2 Background

2.1 Digital hate culture

2.1.1 A hotchpotch of subcultures

It is a shared politics of negation rather than any consistent ideology that unites those spreading hate online: a rejection of liberalism, egalitarianism, 'political correctness,' globalism, multiculturalism, cosmopolitanism, and feminism (Ganesh, 2018; Marwick & Lewis, 2017). Instead of trying to pinpoint a clear-cut culprit, I will use the term *digital hate culture*, as suggested by Ganesh (2018), to refer to a "complex swarm of users that form contingent alliances to contest contemporary political culture" (p. 31), seeking to inject their ideologies into new spaces, change cultural norms, and shape public debate.[2] Marwick and Lewis (2017) speak of "an amalgam of conspiracy theorists, techno-libertarians, white nationalists, Men's Rights advocates, trolls, anti-feminists, anti-immigration activists, and bored young people [who] leverage both the techniques of participatory culture and the affordances of social media to spread their various beliefs" (p. 3). Exponents of digital hate culture can roughly be divided into two groups, ideologues and so-called 'trolls'.

The slippery term 'trolling' encompasses a wide range of asocial online behaviors: it can subsume relatively innocuous pranks and mischievous activities where the intent is not necessarily to cause distress, but it can also take more serious forms of online behavior through which trolls seek to ruin their target's reputation, reveal sensitive or personal information ('doxing'), or spread misinformation (Ibid.). Trolls aim to provoke emotional responses, a behavior known as the 'lulz' which means to find "humor (or LOLs) in sowing discord and causing reactions" (Ibid., p. 4). By claiming their actions to be apolitical and merely 'for fun,' trolling provides "violent bigots, antagonists, and manipulators a built-in defense of plausible deniability, summarized by the justification 'I was just trolling'" (Phillips, 2018, p. 4). This justification allows users to spread shockingly racist, sexist, and xenophobic content without having to fear legal persecution.

---

[2] I refer to the concept of digital hate culture as "linguistically, spatially, and culturally bound" to social media activity in Europe and the US (Ganesh, 2018, p. 32). Although similar dynamics are evolving throughout the world (Ibid.), it is not feasible to account for all specificities and deviations within the scope of this thesis.

The group of ideologues – those exponents of digital hate culture who openly display and admit to a political agenda – cover a wide spectrum of right-wing positions with differing levels of extremism. According to Teitelbaum (2017), we can understand today's right-wing spectrum in three main dimensions, comprising i) race revolutionaries (fringe groups of white supremacists and National Socialists) who celebrate the cause of historical Nazism and declare national identity to be a matter of blood, ii) cultural nationalists (the mainstream populist right) who claim that national identity derives from cultural practices and allege that Islam poses the greatest threat to Western societies, and iii) Identitarians (the European *Identitarian Movement* and the US *alt-right*) who understand race and ethnicity as indispensable elements of identity but – in contrast to race revolutionaries who proclaim the superiority of the white race – purport not to oppose any particular ethnic or religious 'other' but merely strive to 'protect' the true diversity of mankind by propagating against immigration (Ibid., p. 4f.). While race revolutionaries have long used the Internet in the hope of recruiting new members (Marwick & Lewis, 2017), the relatively recent emergence of the *Identitarian Movement* and *alt-right* expanded the online presence of right-wing ideology significantly.

Despite their different motivations, levels of extremism, and choices of antagonists, many of these groups converge on common issues – most significantly the fight against immigration and multiculturalism (Teitelbaum, 2017) – and all of them share their embeddedness and appropriation of Internet culture (Marwick & Lewis, 2017). They take advantage of the "current media ecosystem to manipulate news frames, set agendas, and propagate ideas" (Ibid., p. 1). There is a wide range of blogs, forums, podcasts, image boards, chatrooms, and social media platforms that attract exponents of digital hate culture with lax community policies and strong advocacy of free speech, also referred to as 'Alt-Tech platforms' (Ebner, 2019).[3] However, online hatred also finds expression in mainstream social media platforms (Figure 1). There, it becomes harder, if not impossible, to determine intentions, clear ideological affiliations, and the level of organization of users spreading hate and misinformation. Thus, it is crucial to recognize the practices that digital hate culture operates through.

---

[3] The most well-known ones of them being 4Chan, 8Chan, Reddit, Gab, VK, and Pewtube.
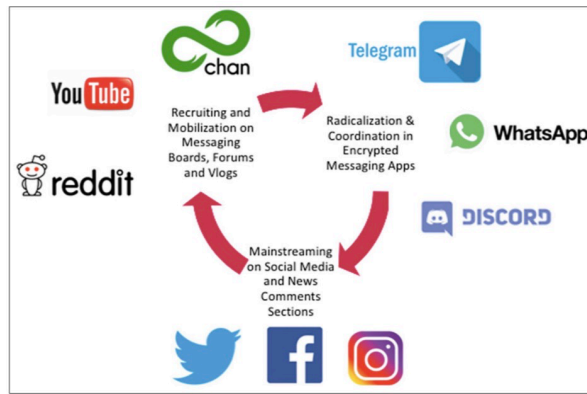
Figure 1: Online influence ecosystem of digital hate culture (Ebner, 2019, p. 172)

2.1.2 Denigration and manipulation practices

Online hate speech[4] is undoubtedly the pivotal form of expression of digital hate culture. A survey in Germany from 2018 shows that 78 % of Internet users had already been confronted with hate speech online, among the 14 to 24-year-old users it is 96 % (Landesanstalt für Medien NRW, 2018). The phenomenon of hate speech has undergone different definitions and demarcations to terms that are often used in the same context, such as cyberhate, cyberbullying, and dangerous speech. At the core of existing definitions lays the emphasis of an identity-based enmity towards certain groups. Although hate speech typically exploits a fear of the unknown based on identity categories like race, ethnicity, religion, and gender (Leonhard, et al., 2018), this is not always the case. The scapegoating can also be directed against the establishment, its institutions like the mainstream media (*Lügenpresse*) and its proponents (*Systemlinge*), left intellectuals like journalists and academics, and general 'social justice warriors' (*Gutmenschen*) (Darmstadt, et al., 2019, p. 160).

Hate speech comprises the explicit display of textual or visual expressions aiming to deprecate and denigrate (Ebner, 2019), making it a form of norm-transgressing communication (Rieger, et al., 2018) that is "fueled by an ideology of inequality, directed against the democratic principle that all people are created equal" (Darmstadt, et al., 2019, p. 157). While the immediate target of hate speech may also be a single person, "the harm caused […] can extend to entire communities by promoting discrimination and

---

[4] Hereafter, the shorter-term *hate speech* will refer to hate speech in an online environment.

intolerance" (George, 2015, p. 305).[5] Hate speech and its persuasive potential (George, 2015) contribute "to alter social cohesion and democracy by promoting and advocating violence between social groups" (Blaya, 2018, p. 2). Because of its violent potential, hate speech has also been referred to as a form of *dangerous speech* (Schieb & Preuss, 2018). That hate speech indeed carries the risk of promoting offline physical violence has been shown in recent studies. Müller and Schwarz (2017) found a significant correlation between rising hate speech on social media and physical violence towards newly arrived Muslim immigrants in Germany. A similar correlation between the surge of hate crimes against Muslims and the rise of digital hate culture has been supported by a study in the UK (Awan & Zempi, 2017). The real-life threat that denigrative and manipulative practices online pose to pluralist democracies confirm the need for intensified research on the workings of digital hate culture.

Apart from understanding hate speech in its textual form, the deep embeddedness of cyberhate in today's Internet culture also requires to look at the exploitation of visual and aesthetic means that constitutes more subtle and persuasive ways to promote ideology (Bogerts & Fielitz, 2019). In digital hate culture, this becomes most apparent in the wide appropriation of memes. Memes are images "that quickly convey humor or political thought, meant to be shared on social media" (Marwick & Lewis, 2017, p. 36). A meme can be understood as a piece of culture "that passes along from person to person, yet gradually scales into a shared social phenomenon" (Shifman, 2013, p. 364f.). Recent far-right cultures have co-opted this online culture as a gateway to express their radical ideas in an easily digestible and 'funny' way which often facilitates the circumvention of censorship (Bogerts & Fielitz, 2019).

A key manipulation practice that often comes along with hate speech, is the spread of misinformation or so-called 'fake news'.[6] 'Fake news' has become a buzzword to describe a wide range of disinformation, misinformation, and conspiracy theories circulating online and in the media (Marwick & Lewis, 2017). In most cases, it is the willful intention that "differentiates fake news from sloppily researched reporting"

---

[5] In delimitation to hate speech or *cyberhate* (Blaya, 2018) this, *cyberbullying* (Ibid.) is a form of online hate that is "characterized by the derogation and defamation of single individuals" (Rieger, et al., 2018, p. 461) without resulting in greater consequences on whole communities.

[6] The term has been broadly circulated since Donald Trump was elected President of the United States in 2016. Initially, it was used to describe websites "that intentionally posted fictional partisan content as clickbait, but Donald Trump's administration quickly adopted it to discredit accurate but unflattering news items, ironically making the term itself a form of disinformation" (Marwick & Lewis, 2017, p. 44).

(Darmstadt, et al., 2019, p. 162). Typically, exponents of digital hate culture berate established media as 'fake news' or 'lying press' and consult alternative news sources that support the own 'truthful' world-view (Marwick & Lewis, 2017, p. 44). Today's historically low levels of trust in mainstream media is a self-perpetuating phenomenon.

> "Groups that are already cynical of the media – trolls, ideologues, and conspiracy theorists – are often the ones drawn to manipulate it. If they are able to successfully use the media to cover a story or push an agenda, it undermines the media's credibility on other issues. Meanwhile, more members of the public are exposed to its weaknesses and may turn to alternative news sources" (Marwick & Lewis, 2017, p. 45).

The different manifestations of these denigrative and manipulative practices, as well as their effects, are crucial to the analysis of both the user discussions on Facebook and the production practices of the case that serves to exemplify digital hate culture in this thesis, the German 'troll network' *Reconquista Germanica* (RG).

2.1.3 Reconquista Germanica[7]

Calling itself the 'biggest patriotic German-language Discord[8] channel' (Davey & Ebner, 2017, p. 19), RG's primary goal is to reclaim cyberspace, influence online discourse, and exert pressure on politics (Kreißel, et al., 2018). The case of RG demonstrates that trolling is not only a practice of apolitical gamers and nerd subcultures who are after the 'lulz,' but that it also attracts and subsumes internet users with a political agenda and ideological motivation. RG, which claims to be a mere "satirical project that has no connection to the real world" (Anders, 2017), is believed to have played a crucial role during the German parliamentary elections in 2017. Inspired by the US *alt-right*'s online engagement that pushed forward the election of Trump, the German far-right joined forces striving to influence the outcome of the German election in favor of the right-wing populist party *AfD* (Kreißel, et al., 2018). The group had almost 500 members on the first day after its founding and on the day before the general election, RG counted about 5000 members willing to invade the Internet with coordinated raids to polarize and distort public discourse (Kreißel, et al., 2018; Bogerts & Fielitz, 2019). The Discord server was set up

---

[7] The term 'reconquista' refers to the reconquest of the Iberian peninsula from the Muslim Moors in the end of the 15th century and represents a historical analogy to the current situation in which far-right groups believe to face a Muslim invasion that needs to be fought back (Kreißel, et al., 2018).
[8] Discord is an encrypted gaming application.

by a German far-right activist under the pseudonym Nikolai Alexander, whose Youtube channel – loaded with conspiracy theories and anti-Semitism – counted over 33.000 subscribers before it was blocked (Davey & Ebner, 2017). With the RG Discord server, Nikolai Alexander built a far-right 'troll army,' assembling 'patriotic forces' from a wide spectrum of cultural nationalists, loosely organized Neo-Nazis, far-right influencers, the Austrian and German sections of the *Identitarian Movement*, and members of the right-wing parties *AfD*, *Junge Alternative*, and *NPD* (Bogerts & Fielitz, 2019; Ayyadi, 2018).

The group is managed based on strict hierarchies appropriating a mixture of military language derived from the Nazi-German *Wehrmacht* and *alt-right* insider vocabulary (Davey & Ebner, 2017, p. 19) with Nikolai Alexander being the self-appointed commander-in-chief (Figure 2, 3). To become a member or to move up to higher 'army ranks,' users have to pass special recruiting procedures and prove commitment by joining organized raids (Bogerts & Fielitz, 2019, p. 140). Racist, xenophobic, and National Socialist content and symbols are so dominant in the group (Kreißel, et al., 2018) that it seems surprising that the German Federal Government has declared that RG does not classify as a right-wing extremist network and is hence not subject to observation for the *Federal Office for the Protection of the Constitution* (Deutscher Bundestag, 2019).

Figure 2: Meme from RG's Discord server (Ayyadi, 2018)

Figure 3: RG's server hierarchy (Ayyadi, 2018)

2.2 Responses to digital hate culture

2.2.1 Regulations and legal actions

Policymakers and legal actors are aware of the danger digital hate culture poses to society and propose regulatory measures. One attempt to oppose cyberhate has been the German *Network Enforcement Act* (short NetzDG). It forces social media companies to implement procedures that allow users to report illegal or extremist content and, crucially, it obliges the platforms to canvass these user complaints immediately and to delete content "obviously punishable by law" within 24 hours or seven days in less clear incidents. Moreover, the social media companies are obligated to give account for their resources, teams, and measures dedicated to deleting hateful content (Darmstadt, et al., 2019). Such transparency of curating practices is a clear benefit of the legislation and from an international perspective, the law constitutes a novel and aggressive approach to crack down on digital hate culture (Ibid.).[9] Procedures to repress content, as suggested by NetzDG, have, however, also gained substantial critique as they raise difficult questions about the limitations of free speech and the definition of where hate begins and what should be deleted (Rieger, et al., 2018). Depending on what perspective is weighted most – the sender's, the receiver's, or the observer's – the evaluation of malicious content can vary substantially (Ibid.). Although the law contains a whole catalog with criminal acts that count as hate speech[10], "real life tends to complicate its applicability as an underwriter for anything like legal action" (Thurston, 2019, p. 194). Moreover, such legislations shift the responsibility to decide over the legality of content to the employees of private companies, which has been criticized by experts and NGOs as "unfortunate privatization of law enforcement" (Ibid.). The impending high fines in case of infringement are likely to lead to an 'overblocking' of content that is not unlawful. For the users themselves, it is mainly the conflict between NetzDG and §5 of the German Basic Law protecting the freedom of speech that evokes strong opposition. The sentiment

---

[9] In the first seven months after the law went into effect on January 1 2018, more than 500.000 posts (on Facebook, Twitter, and Google) were flagged as being inappropriate (Gollatz, et al., 2018).
[10] As Jaki and De Smedt (2018) explain, "[u]nder German criminal law, illegal forms of hate speech include incitement to criminal behavior (*Öffentliche Aufforderung zu Straftaten*, §111 StGB), incitement of the masses (*Volksverhetzung*, §130 StGB), insult (*Beleidigung*, §185 StGB), defamation (*Üble Nachrede*, §186 StGB), slander (*Verleumdung*, §187 StGB), coercion (*Nötigung*, §240 StGB), […] intimidation (*Bedrohung*, §241 StGB) […] [and] defamation or slander of a politician (*Üble Nachrede und Verleumdung gegen Personen des politischen Lebens*, §188 StGB)" (p. 6).

to perceive and frame the law as a form of censorship can ultimately fuel radicalization itself (George, 2016).

Crucially, such legislations miss the point that most hate speech and the spread of false information is not punishable by law and exponents of digital hate culture often operate *strategically* meaning that they are well aware of legal boundaries and apply measures to circumvent them. Most ideologues and trolls are clever enough not to explicitly deny the Holocaust or spread Swastikas. It is, however, not illegal to invent and spread lies. Threats are carefully formulated, humor and hidden references are utilized, and operations are relocated to 'Alt-Tech platforms'. For these reasons, it is not surprising that many actors call for alternative strategies to fight digital hate culture, such as the promotion of counter communication which this thesis focuses on as one possible response.

2.2.2 Counter communication

Think tanks, NGOs, and civil society initiatives have been looking for ways to meet hate speech with counter speech. Although scholarly definitions on 'counter speech' are scarce, there are aspects that have repeatedly been emphasized. As hate speech's antagonist (Richards & Calvert, 2000), counter speech encompasses all communicative actions aimed at refuting hate speech and supporting targets and fellow counter speakers through thoughtful and cogent reasons, and true fact-bound arguments (Schieb & Preuss, 2018). Counter speech may further have preventative effects as it "can disseminate messages of tolerance and civility, contributing to civic education in a broader sense" (Rieger, et al., 2018, p. 464). Although counter speech efforts are sometimes regarded as the currently most important response to hate speech (Schieb & Preuss, 2018), such measures have also received critique. Undoubtedly, unwanted side effects can be provoked, when, for instance, counter speakers do not "hold up to civic, participatory standards themselves but act in an authoritative manner or mock about those falling for extremist ideologies" (Rieger, et al., 2018, p. 465). Moreover, counter speech is not always safe, as "speaking out against online aggression can lead to threats and very negative experiences" (Blaya, 2018, p. 7). Nevertheless, there are a number of counter speech initiatives that have successfully taken on the fight against cyberhate and attract a growing number of fellow combatants (Ebner, 2019).

2.2.3 Reconquista Internet

One of the biggest civil society counter speech initiatives in Germany is the group *Reconquista Internet* (RI). It was initiated by Jan Böhmermann, satirist and moderator of the popular German satire news show *Neo Magazin Royal*, at the end of April 2018. The group was initially a satirical reaction to the prior disclosure of RG (Figure 4). Böhmermann's call for a 'love army' to fight back the trolls and reconquer the Internet attracted a large number of supporters, counting some 60.000 people today. RI is a digital, non-partisan, and independent civil rights movement that describes itself as "voluntary and private association of people of all ages, genders, different backgrounds and with all sorts of other individual benefits and mistakes" (Reconquista Internet, 2018). Everyone who respects the German Basic Law and their 10-point codex can join the network.

Figure 4: The launch of RI by Jan Böhmermann (Neo Magazin Royal, 2018)

RI stands up for more reason, respect, and humanity online and aims to "help each and everyone out of the spiral of hatred and disinhibition" (Reconquista Internet, 2018). Just as their opponents, the activists of RI organize their actions on Discord. Their main effort is to seek out online hate speech, mobilize collective action, and refute hateful content by means of counter speech. In different 'workshop' areas on Discord, members engage in creating a meme library, develop strategies of counter argumentation, and plan offline and online actions.

3 Literature Review

3.1 Online hatred and far-right communication practices

With digital hate culture proliferating over the Internet and the calls for effective responses becoming louder, research in the field has stepped up lately. Studies investigate the content and strategic operating of cyberhate, based on which researchers advise means of dealing with the destructive dynamics. And while only a few years ago, digital communities were irrelevant for those studying far-right communication, today, online spaces are an indispensable part of the research agenda on antagonistic politics (Bogerts & Fielitz, 2019, p. 138).

Aiming to "give an insight into what disparaging verbal behavior from extremist right-wing users looks like, who is targeted, and how" (p. 1), Jaki and De Smedt (2018) analyzed over 55.000 German hate tweets during a period of nine months in 2017/2018. With a qualitative and quantitative analysis of hate speech from a linguistic perspective, the authors found that immigrants, and specifically Muslim refugees, are the group most targeted by online hatred in Germany, which supports the topical focus chosen for this thesis. Their research shows that hate speech is marked by some predominant types of persuasive speech acts, including expressive speech (aggressive speech often accompanied by emojis), directive speech (calls for action often accompanied by hashtags) as well as assertive, commissive, and indirect speech (Ibid., p. 10f.). Conforming with Jaki and De Smedt's research, Davey and Ebner (2017) observe that the hatred towards Muslims is the pivotal nexus point across right-wing groups. In a three-month ethnographic research project, the authors mapped the ecosystem of the 'new' far-right across Europe and the US analyzing content gathered from over 50 different platforms. One of their key findings is that "[e]xtreme-right groups across the globe are actively collaborating to achieve common goals, such as keeping refugees out of Europe, removing hate speech laws and getting far-right populist politicians to power" (Ibid., p. 5) by appropriating shared online communication strategies, including "in-group jokes and memes, cultural references and military vocabulary" (Ibid., p. 28). Another strategy pursued by exponents of digital hate culture has been explained to lay in the purposeful use of rumors and false reports to create what Darmstadt et al. (2019) call *toxic narratives*. It is characteristic for far-right echo chambers, the authors explain, to interweave several

de facto unconnected narratives and spread the toxic narratives that are established this way. In their case study, the authors demonstrated how the murder of a 14-year-old girl in Berlin in 2018 was linked to other murders of girls by Muslim men which established an insider reference "that creates assumed connections between refugees, Islam, and violence" (Ibid., p. 159).

Next to these studies investigating the practices of digital hate culture, there are also scholars who have critically assessed how the platforms on which hatred is disseminated facilitate and benefit the cause of cyberhate (e.g., Matamoros-Fernández, 2017 & Massanari, 2017). In a study on the material politics of Facebook, Matamoros-Fernández (2017) illustrates that social media platforms do "not only host public communication, but also coordinate knowledge through their technological affordances and under their logics and rules" (p. 933). She highlights how social media platforms can contribute to circulate overt and covert hate speech through their affordances, policies, algorithms, and corporate decisions (Ibid.). Matamoros-Fernández' concept of 'platformed racism,' which she understands as a new form of racism derived from the material politics of social media platforms, allows shedding light on the role that social media platforms play as spaces that facilitate 'platformed hatred'.

Departing from these previous insights into the workings of digital hate culture, this thesis seeks to further contribute to the understanding of cyberhate's destructive nature on mainstream social media. The multi-perspectival approach makes this study stand out from previous research efforts: this thesis considers the intended practices of an organized group of digital hate culture as well as the actual manifestation of the practices of online hatred visible on Facebook and seeks to bring them into dialogue.

3.2 The effects and potential of counter speech

Compared to the upsurge in studies on digital hate culture, research on counter efforts is still in its early days. Blaya (2018) conducted a review of intervention programs to cyberhate and emphasized the lack of rigorous assessment. She stresses the need for intensified research in the field, especially on the effectiveness of interventions to prevent and reduce cyberhate. However, in practice, it has proven difficult to "measure" the constructive impact of counter speech as Silverman et al. (2016) underline. Hence, there

is little consensus around what works and what does not. Some scholars have nevertheless attempted to assess the effectiveness of counteraction.

Schieb and Preuss (2018) assessed the persuasive effect of counter speech. Through an experiment with four modeled user groups including the *core* ('haters'), *clowns* (trolls), *followers* (the non-decided audience), and *counter speakers*, the authors find that i) depending on the situation, counter speech should not take too extreme positions but be rather moderate to achieve maximum impact, ii) the first speaker always has an advantage moving the average opinion to her/his side, and iii) counter speech works best if it is organized, conducted in groups, and quick in reaction (Ibid.). Investigating the potential of counter speech on Facebook, Miškolci et al. (2018) explored the impact of two specific counter strategies, that is, fact-checking (refuting prejudices and myths through fact-based information) and personal experience (highlighting positive experiences with the targeted group). By looking at reactions to hate speech posts directed against Roma in Slovakia, the authors analyzed 60 Facebook discussions on Roma-related topics. The qualitative content analysis revealed a prevalence of anti-Roma attitudes but demonstrates that "pro-Roma comments encouraged other participants with a pro-Roma attitude to become involved" (Ibid., p. 1). Counter speech strategies, especially fact-checking, were shown to have their limitations. The results highlight that people tend to overestimate their knowledge of common social phenomena and, instead of reconsidering their opinion based on scientific data, they refuse evidence thinking they already have enough information. Despite attempts to counter hate speech on Facebook, users, therefore, most often continue in their hate (Ibid.). Sponholz (2016) comes to a similar critical assessment of the potential of counter speech. With a focus on anti-Muslim discourse in Italy, she investigated journalistic responses to the newspaper publication of an Islamophobic pamphlet by the author Oriana Fallaci. Through a quantitative content analysis of the counter-responses to the anti-Muslim sentiment, Sponholz found that "counter-speech did not lead either to consensus or to refutation of such contents" (Ibid., p. 502) but instead improved hate speech by "providing Islamophobia with relevance, legitimacy ('discussability') and better discourse quality" (Ibid., p. 517).

Neumayer and Valtysson's (2013) research is one example of how counter speech has been studied through a multidimensional framework that pays attention to the functionalities of social media platforms. Based on three case studies on nationalist

demonstrations in Germany, the authors conducted a qualitative content analysis of tweets, asking how power is reproduced and challenged in anti-fascist counter-protests on Twitter (Ibid., p. 4). Neumayer and Valtysson identified tactics and strategies that activists use to both contest and reproduce power in interplay with the functionalities of Twitter, such as, for instance, the sharing of values through hashtags and retweets, mutual monitoring through observing hashtags, and making use of spectacle, play, and humor (Ibid.). The authors outline how existing imagined communities become constituted as networked publics through the functionalities of the social media platform. With theoretical reference to Laclau and Mouffe (1985), Neumayer and Valtysson explain how networked publics "overcome differences within 'democratic pluralism' […] to form a collective against a 'common enemy'" (Ibid., p. 18). Another study drawing on Mouffe's political theory to investigate hateful content online has been conducted by Milner (2013). He assessed the potential of the 'logic of lulz' to be adopted by counter publics and afford adversarial pluralism over exclusionary antagonism. With a focus on memes, Milner undertakes a critical discourse analysis of discussions on the participatory media collectives 4Chan and Reddit. He finds that the irony-laden communicative practices underlying the 'logic of lulz' are not always essentializing marginalized others but can also be appropriated by counter publics to foster a vibrant, agonistic discussion (Ibid., p. 62).

Similar to these previous efforts of researching and assessing counter communication, this thesis seeks to highlight the practices, potentialities, and limitations of counter speech aiming to confront digital hate culture. Again, the uniqueness of this approach is that the analysis of strategic communication documents extends the scope of research to allow reflections on the intended practices as well as on the realization of such at the place of action. Moreover, the analysis takes a holistic multidimensional approach that goes beyond looking at practices only but also considers the underlying structures deriving from technological affordances, power relations, and discourses that crucially determine the workings of cyberhate and counter speech.

4 Theoretical Foundations and Analytical Framework

4.1 'Agonistic pluralism' and the post-political Zeitgeist

The work of Chantal Mouffe on today's post-political *Zeitgeist* and *the political* provides intriguing perspectives that can help to understand the anti-democratic positions and behaviors flourishing in digital hate culture and allows to generate impetus for how to deal with the poisoning of public discourse without jeopardizing an already fragile democracy. In *On the Political* (2005), Mouffe states the foundations of her understanding of politics. She holds the post-structuralist view that identities are always relational and that the nature of collective identities always entails a 'we/they' discrimination (p. 5). With critical reference to Carl Schmitt's emphasis on the conflictual nature of politics, Mouffe develops an understanding of *the political* as an ever-present possibility of a 'friend/enemy' relation (Ibid., p. 15). She considers such antagonism to be constitutive of human societies (Ibid., p. 9). In this eternal sphere of conflict, it is the role of democratic *politics* to create order. According to Mouffe, the crucial challenge for any democracy is to facilitate and foster a 'we/they' relation that allows 'antagonism' to be transformed into 'agonism'.

> "While antagonism is a we/they relation in which the two sides are enemies who do not share any common ground, agonism is a we/they relation where the conflicting parties, although acknowledging that there is no rational solution to their conflict, nevertheless recognize the legitimacy of their opponents. They are 'adversaries' not enemies." (Ibid, p. 20)

Her argument for agonistic politics essentially means to keep the 'we/they' relation from devolving into a 'friend/enemy' relation and instead to advance an understanding of the opposing positions as 'adversaries'. Therefore, Mouffe sees the main task of democracy in providing space for a plurality of adversarial positions and ideological conflicts and in ensuring them legitimate forms of expression (Ibid., p. 20f.). She accentuates the positive role of the parliamentary system to facilitate partisan conflict and cautions against the contemporary post-political vision that seeks to go beyond a 'left and right' distinction.

Mouffe challenges the *Zeitgeist* of the post-political world and the hegemony of liberalism that developed after the Cold War and with the advent of globalization. She sees liberalism's central deficiency in its "negation of the ineradicable character of antagonism" (Ibid., p. 10). Liberal thought, characterized by methodological

individualism, does not acknowledge the nature of collective identities, making it unable to "adequately grasp the pluralistic nature of the social world" (Ibid.), including the conflicts that pluralism entails.[11] Such conflicts cannot be resolved by recourse to universal human values, liberal consensus, and rationality, Mouffe argues. In fact, by refusing to acknowledge the antagonistic dimension constitutive of *the political*, the consensual approach and the promotion of a world beyond 'left and right' and 'we and they' advocates an 'anti-political vision' (Ibid., p. 2) that only exacerbates the antagonistic 'friend/enemy' distinction present in today's society. Without a real choice between significantly different political parties, disaffection with politics sets in, nurturing the formation of other types of collective identities around nationalist or religious forms of identification (Ibid., p. 30). Antagonisms, she explains, "can take many forms and it is illusionary to believe that they could ever be eradicated" (p. 30), making it crucial "to allow them an agonistic form of expression through the pluralist democratic system" (Ibid.). In this context, Mouffe explains the rise of right-wing populism across Europe as a consequence of the blurring frontiers between left and right and the lack of vibrant partisan debate (Ibid., p. 70). Politics, she ascertains, is increasingly played out in a moral register, in which the 'we/they' – 'the people' vs. 'the establishment' – turns into a dichotomy of the 'evil extreme right' and the 'good democrats' instead of being defined with political categories (Ibid., p. 73). Her argument is that "when the channels are not available through which conflicts could take an 'agonistic' form, those conflicts tend to emerge on the antagonistic mode" (Ibid., p. 5). Instead of trying to negotiate a compromise or rational (i.e. a fully inclusive consensus) solution among the competing interests, Mouffe proposes a radicalization of modern democracy, advocating 'agonistic pluralism,' in which conflicts are provided with a legitimate form of expression (Ibid., p. 14).

Several scholars have embedded their research on the antagonistic politics of digital hate culture in the theory of Mouffe (e.g., Cammaerts, 2009; Milner, 2013; Neumayer & Valtysson, 2013; Davids, 2018). While Milner (2013) and Neumayer and Valtysson (2013), as outlined in chapter 3.2, focus on the potentialities of Mouffe's theory to highlight the possibilities of agonistic pluralism, Cammaerts (2009) and Davids (2018) raise critical questions about the limitations of agonistic pluralism. Cammaerts (2009)

---

[11] Mouffe elaborates on the optimistic liberal understanding of pluralism which suggests that "we live in a world in which there are indeed many perspectives and values and that […] we will never be able to adopt them all, but that, when put together, they constitute an harmonious and non-conflictual ensemble" (2005, p. 10).

shares Mouffe's critique on the optimistic Habermasian understanding of the Internet as a public sphere for political discussion and consensus-making. In his case study on north-Belgium far-right discourses on blogs and forums, he highlights how the clash of political positions online is an expression of the antagonistic nature of politics: "ideologically heterogeneous unmoderated spaces for debate, while being more open, are often confronted with flame-wars between (often anonymous) participants" (Ibid., p. 4). Although Cammaerts adopts Mouffe's perspective by considering how such confrontations can be seen as beneficial for a strong democracy, he also asks the important question "whether being a racist is a democratic right [and] whether freedom of speech includes opinions and views that challenge basic democratic values" (Ibid., p. 1). In doing so, he calls attention to the particularities that need to be considered when drawing Mouffe's theory into the hate speech debate, namely to reflect on how absolute freedom of speech and the nature of democracy is and how far the right to respect and recognition of ideological difference and the right not to be discriminated against goes (Ibid., p. 2). Davids (2018) expresses similar concerns when discussing the extent to which democratic tolerance can offer a plausible response to hate speech. She argues that it is the act of attacking the morality in others, rather than attacking political standpoints, that makes hate speech intolerable in a democracy.

This thesis follows Mouffe's understanding of politics as a battlefield where contesting groups with opposing interests vie for hegemony. The clash between exponents of digital hate culture and counter speakers online showcases the antagonism flourishing in contemporary society. With the analysis of such confrontations, this thesis takes up Mouffe's theory to discuss the agonistic potential of discussions on social media, considering the conditions under which the conflict is played out and reflecting on the limits of legitimate positions of adversaries in a pluralist democracy.

4.2 A four-dimensional approach

To present a more holistic picture of the interplay between digital hate culture and counter speech efforts on social media, the analysis is guided by a four-dimensional framework suggested by Uldam and Kaun (2019) for studying civic engagement on social media. The approach avoids a techno-deterministic and media-centric focus and instead

considers the context. Covering each of the four dimensions with the proposed research questions, the analytical framework allows for a goal-oriented analysis that focusses on intended and actualized (i) practices (RQ 1) as well as the embedded (ii) discourses (RQ 2) but also provides space to consider the role of (iii) power relations and (iv) technological affordances (RQ 3) as well as their interrelatedness (Figure 5).
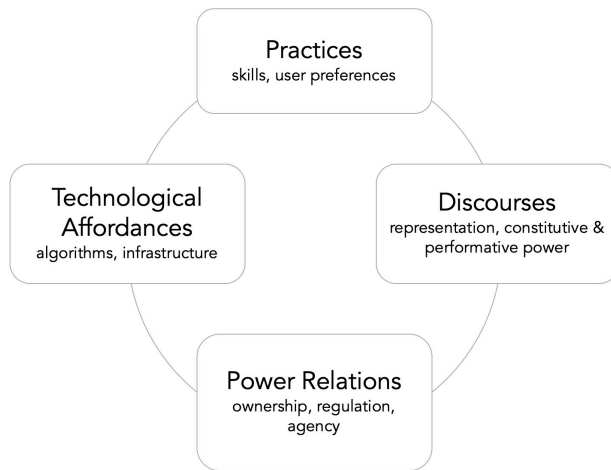

Figure 5: An analytical model of civic engagement on social media (adopted from Uldam and Kaun (2019, p. 104))

The pyramid can be turned four ways up, placing the apex at the top that is prioritized in a research project (Ibid., p. 103). As it is not feasible to pay equal attention to all four focal points within the scope of this thesis, the analysis is focused on the dimensions practices (RQ 1) and discourses (RQ 2) to which both the theoretical and methodological framework are aligned to. Placed on top of the pyramid, the dimension *practices* concentrates on what people *do* when using social media to spread hatred or counter-argue and *how* they do it. Building on the findings that the empirical material reveals in this regard, the analysis then turns to consider the role of discourses. *Discourses* condition our understanding of the world and permanently compete over what is considered 'true'. Representations online contribute to form and perpetuate discourses and are hence a crucial part of this struggle over the creation of meaning. Looking at discourses thus allows acknowledging the constitutive and performative power of user interactions. Although closely connected, *power relations* focus more on questions of ownership and regulation, shedding light on dynamics of privilege and (non-)visibility and the ethical implications of such. Promoting such power hierarchies, a glance at *technological*

affordances underlines how Facebook's algorithms affect the interplay of hatred and counteraction. As such underlying structures and other actors (e.g., page hosts and social media companies) essentially constitute the context in which any user interaction on social media is situated, I consider it crucial to include a discussion of power relations and technological affordances (RQ 3); not to analyze their manifestation and impact in sufficient depth but mainly to touch upon the importance of social media research to pay attention to the contextual setting of user interaction which is by no means neutral.

5 Methodology

5.1 Choice of method: qualitative content analysis

Similar to some of the studies presented in the literature review (Miškolci, et al., 2018; Ernst, et al., 2017; Neumayer & Valtysson, 2013), the empirical material of this research project is analyzed with the help of qualitative content analysis. The focus thereby lies on detecting themes and patterns that illustrate the range of meanings of a research phenomenon. To meet the demands of scientific rigor, the description and subjective interpretation of textual data follow the systematic classification process of coding. As the main objective of this research is to elicit meaning in the four dimensions discussed above, the qualitative content analysis serves as an initial analytical step to organize the corpus of data before engaging in the more essential process of interpreting and contextualizing the data. Zhang and Wildemuth (2005) put forth an eight-step model for processing data that is based on systematic and transparent procedures. My analysis follows the single steps of this model and simultaneously draws on techniques and basic principles put forth by Mayring (2000, 2014). For the first dataset, the strategic communication documents, the analysis follows Mayring's (2014) *summary of content*, which helps to reduce the material to its core content by paraphrasing und generalizing the data. The second dataset, the user comments, is analyzed through *structuring of content*, in which aspects of the material are filtered out based on pre-determined ordering criteria. The method is tailored for an analysis of the empirical material within the two focus dimensions, practices and discourses. Departing from the findings of this analysis, the discussion then turns to discuss the role of technological affordances and power relations. This additional contextualization is based partly on indicators found in the data,

and partly on textual sources including previous research efforts and Facebook's community directives.

5.2 Sample and strategy of data collection

5.2.1 Strategic communication documents

As samples in qualitative research often consist of purposively selected texts that best inform the investigated research questions, I followed non-random sampling strategies to select material that appears suitable to shed light on the practices and discourses involved in the interaction between exponents of digital hate culture and counter speakers. This sampling resulted in two datasets: user comments from discussion threads on Facebook and selected strategic communication documents of RG and RI. For the latter, I chose all documents of both RG and RI that deal with concrete media strategies and tactics of communication advocated by the groups and that were accessible publicly or with the consent of the respective group. In the case of RG, the sample consists of the 10-page document *Handbook for Media Guerillas* which is composed of four parts providing how-to guidelines for 1) shitposting, 2) open-source memetic warfare, 3) social networking raids, and 4) attacking filter bubbles (D-Generation, 2018). *HoGeSatzbau* ('Hooligans Against Syntax'), an initiative that counteracts nationalist content, leaked the document from RG's Discord server and made it public in January 2018 (Lauer, 2018).[12] Having public access to these strategic documents that are otherwise sealed off in a closed group offers a unique opportunity to gain insights into the organized operating and intended practices of digital hate culture. In case of RI, the sample comprises the group's wiki page on Reddit, in which all project relevant information as well as the group's 10-point-codex, guides, and instructions are compiled (Reconquista Internet, 2019).

---

[12] Originally, the document was published on the website of the far-right group *D-Generation*, where it has already been circulated in May 2017 (D-Generation, 2017).

5.2.2 User comments

To sample user comments on Facebook, a set of articles from mainstream journalistic sources with relevant discussion threads had to be selected. In order to do so, I followed the counter speech actions of RI. Once verified for the group's Discord server, I gained access to different workshop areas, including *WS Counterspeech*. There, members post links to discussions in different social media platforms where they have detected hateful content and call for fellow members to counteract. The sub-chat for Facebook was an effective means to identify a sample of comment threads that are likely to showcase an active involvement of organized counter speakers as well as an engagement of exponents of digital hate culture. As the review of previous literature on hate speech has shown, anti-Muslim sentiment is currently the most central nexus point for the diverse groups constitutive for digital hate culture internationally (e.g., Jaki & De Smedt, 2018; Davey & Ebner, 2017). Positioning my research among previous efforts to shed light on and undermine this alarming condition, the journalistic articles on Facebook were chosen purposefully based on their thematic relation to Islam or the immigration of Muslims to Germany. Following the links posted by RI members in the period from February to March, the data collection resulted in the following three articles:

TAG24 Dresden 9 hrs ·	In a short news story, the regional news portal *TAG24 Dresden* reports on Claus-Peter Reisch, captain of the rescue ship *Lifeline*, who advocates sea rescue operations at a rally of the humanitarian organization *Mission Lifeline* in Dresden (Tag24 Dresden, 2019).

Migranten in Seenot retten - das ist für "Lifeline"-Kapitän Claus-Peter Reisch (57) eine Herzensangelegenheit.

TAG24.DE
"Mission Lifeline"-Kapitän wirbt für mehr Menschlichkeit
Lifeline-Kapitän Claus-Peter Reisch war am Samstag in Dresden und...

216 · 133 Comments 23 Shares

Published on Facebook 24.02.2019

Table 1: Article I

A commentary by the editor-in-chief of the regional tabloid *B.Z.* assesses the terrorist attack in New Zealand in which a right-wing extremist killed 49 people in two mosques. The article remarks that the crime showed that revenge and hate calls for more revenge and hate, concluding that it does not help anyone to condemn people because others do evil (B.Z., 2019).

Table 2: Article II

The article reports on a study conducted by the *Friedrich Ebert Foundation* which investigated public sentiment on the topic of migration in Germany. Results are presented in data visualizations and explanatory text, showing that a majority of Germans rates immigration as an opportunity (53 %) but also opposes the reception of more refugees (56 %) (Spiegel Online, 2019).

Table 3: Article III

The comment threads under each article were archived within 24 hours after publication (avoiding the deletion of content under the regulations of NetzDG), comprising all user comments (709) and responses to comments in full length. To illustrate the findings of the analysis with examples, selected visuals of comments, conversations, and memes are included in the presentation of results.

5.3 Steps of analysis

First, to prepare the data for analysis, all documents and comment threads were uploaded to NVivo.[13] In a second step, the units of analysis were determined. As coding units in qualitative content analysis are not bound to specific linguistic units but rather represent the established themes (Zhang & Wildemuth, 2005), the analyzed units varied between single words, sentences, paragraphs, or entire user comments. Next, themes in the form of categories were developed, which were derived both inductively from the data itself and deductively from previous studies and theory. Combining these approaches is suitable as it ensures attentiveness to the data's individual properties but also accounts for the extensive research in the field during the last years (Mayring, 2014; Zhang & Wildemuth, 2005). The determined categories were aligned to the two dimensions of the analytical framework the content analysis focuses on, practices and discourses. For the strategic communication documents, the three main categories were *strategies and tactics*, *advocated values*, and *target groups*. The user comments were coded based on the main categories *type of comment* (comprising the codes *digital hate culture*, *counter speech*, *page host*, and *meme/GIF*) and *discourse* (comprising the codes *Islamophobia/anti-refugee*, *anti-establishment/anti-left*, and *nationalism*). Based on coding schemes (Appendix) that hold all the categories and systematize the process of data analysis, the coding was carried out, first on a sample to validate coding consistency (Zhang & Wildemuth, 2005) and subsequently on the entire corpus of text. In the process of coding the data, the coding scheme was continuously complemented with categories that derived inductively from newly discovered themes. After another assessment of coding consistency, I reflected on my methods before deriving meaning and drawing conclusions from the coded data.

---

[13] Programs like NVivo help to organize, manage, and code qualitative data in a more efficient manner (Zhang & Wildemuth, 2005). For this thesis, NVivo assisted the processes of data collection, text editing, note-taking, coding, text retrieval, and the visual presentation of categories and their relationships.

5.4 Methodological reflections

As an interpretive approach, qualitative research differs from the positivist tradition and therefore requires different assessment criteria than the conventional criteria validity, reliability, and objectivity (Zhang & Wildemuth, 2005). Lincoln and Guba (1985) proposed an alternative assessment based on the four criteria credibility, transferability, dependability, and confirmability. To ensure the quality of this research project according to these criteria, I placed special emphasis on providing a detailed account of the procedures for data collection and analysis as well as for decision points in setting up and carrying out the study. In this way, the study can be sufficiently contextualized by future researchers in order to "determine the extent to which their situations match the research context, and, hence, whether findings can be transferred" (Merriam, 2009, p. 229). Being aware of the unique perspective each researcher brings to a qualitative study, I aimed to provide the reader with sufficient descriptions to understand the basis for any interpretation made in the presentation of results. To recognize the subjective voice of the researcher also implies to reflect on and convey my own positionality in the research process. Since I endorse social justice and equality and believe in democratic values, I oppose any form of identity-based discrimination as promoted by exponents of digital hate culture. My clear rejection of Islamophobic sentiment has initially sparked my interest in researching *effective* ways of countering hate speech. Hence, my personal beliefs and moral stances are carrying biases that are likely to affect the process of data collection and analysis. However, by trying to distance myself from an overly optimistic view on the potential of counter speech and by taking a critical stance towards *all* analyzed user comments, I aimed to conscientiously resist the biases of my positionality. I strived to approach the empirical data with an open and curious mindset that provides room for unanticipated observations.

The fact that the research design does not allow to cover all four dimensions of the analytical framework equally must be recognized as a limitation of this research. Considering the scope of this thesis, the role that Facebook's technological affordances and power relations play for the interplay between digital hate culture and counter efforts could not be investigated with the same methodological rigor as the other two dimensions. Another limitation of the research design is that there is no unambiguous connection between the two chosen data sets of strategic communication documents and user

comments. In other words, there is no possibility to be certain that any of the user comments analyzed has actually been produced by either of the two discussed groups RG and RI. The presence of RI's organized counter speakers can be assumed as the link to the user discussions on Facebook has been captured on their internal communication platform, entailing the call for members to become active in the discussions. Although there is no certainty that members of RG participated in the analyzed discussions, the presence of exponents of organized groups of digital hate culture can be assumed based on the high conformity between the analyzed communication strategies of RG and the commenting practices on Facebook, as the presentation of results highlights.

6 Ethics

The research opportunities arising from the vast and easily accessible quantities of data online are always accompanied by a responsibility to ensure that the way in which data is obtained and reused follows the highest possible ethical standards (Townsend & Wallace, 2016). One ethical concern pertains to the question of whether the obtained data should be considered public or private. The strategic communication documents of RI are accessible to anyone through the open online platform Reddit and can thus be considered public data. The *Handbook for Media Guerillas* has been published by several news outlets and was, also before the leak, publicly available on the website of the far-right group *D-Generation*. Thus, all strategic communication documents could be used for research purposes without obtaining informed consent. However, for the process of sample selection, it was essential to become a member of RI's Discord server to gain access to the counter speech chats and use the links shared in these chats for identifying user discussions on Facebook. For using this data, which is only shared with registered members, RI was asked for consent to ensure ethically sound research. As for the second data set, the user comments on Facebook, the extent to which the data must be considered private or public is not as easy to determine. One might argue that all Facebook users have agreed to a set of terms and conditions that contain clauses on how one's data may be utilized by third parties, including researchers, leading to a perception of data as public domain (Ibid.). Users might, however, not be aware of these terms which makes it crucial to evaluate the research ethics for each individual study (Boyd & Crawford, 2012). As the user comments analyzed in this thesis originate from open discussions on mainstream

news media pages and not from private groups or closed discussion forums, I considered it plausible and ethically sound to access and analyze the user comments without seeking informed consent from the users participating in the discussions. Nevertheless, it was imperative to ensure the anonymity and safety of the quoted users. Especially since the data accessed deals with sensitive subject matters, like ideological orientations and xenophobic attitudes, it becomes crucial to protect the identity of unwitting participants to avoid putting any social media users at risk (Thurston, 2019). Social media companies often store data and meta-data for long periods, making comments easily searchable through search engines long after an original posting (Ibid.). Therefore, all user names in the displayed comments were blurred and the quoted comments are only displayed in their translation entailing the necessary alteration of content to ensure that the "data does not lead interested parties to the individual's online profile" (Ibid., p. 11).

Another ethical concern that needs to be addressed touches upon issues of amplification that generally come along with research on harmful or ideological content as such risks benefitting those structures that it strives to undermine. It can be argued that academic research provides harmful groups and their causes with yet another platform for visibility and attention (Askanius, 2019), contributing to amplifying their messages to an even broader public. Providing coherent narratives on their structures and communication strategies and dedicating attention to their activities might affirm exponents of digital hate culture in their own idea of their importance. Although it is crucial to be aware of these risks and reflect on questions as to *why* and *for whom* research is carried out (Ibid.), it remains imperative in contemporary society to continue investigating the anti-democratic workings of digital hate culture and, as it is the case in this thesis, means to counteract their activities. Such research essentially contributes to a knowledge base necessary to undermine and disempower harmful structures, highlighting the moral obligation that researchers carry to engage in studies that encourage positive societal change and provide insights that are of relevance and direct use to counter efforts.

7 Presentation of Results

7.1 Practices

7.1.1 Strategies and tactics of communication

The analysis of the strategic communication documents of both RG and RI reveals how organized forms of digital hate culture and counter initiatives operate systematically to achieve their respective goals. It is apparent that both groups distinguish between two overall communication strategies: one that is tailored for the counterparty – either exponents of digital hate culture or counter speakers – and one that focusses on the audiences, meaning those users who adopt a position somewhere between the two camps. In order to address the oppositional side, in RG's words "the enemy" (D-Generation, 2018), the *Handbook for Media Guerillas* instructs its readers to infiltrate the social media pages of "political parties, […] popular feminists, lackeys of the government […] and all propaganda-government press, like ARD, ZDF, Spiegel and the rest of the fake news clan" (Ibid.). To discredit the "lies" being spread on these pages, the handbook advises users to tag such content as #fakenews and "troll" their opponents with all means possible (Ibid.). The strategy builds upon humiliation, provocation, and discreditation. Specified tactics recommend to create fake accounts in the enemy's name and "post bizarre and preferably stupid content," (Ibid.) to find and exploit the opponent's weak points drawing on a prepared repertoire of insults, to swing the "Nazi club" and turn the enemy's arguments of racism and anti-Semitism against them, to always have the last word, and to pretend calmness and politeness to make the opponent livid with rage (Ibid.).

Contrary to this offensive approach, RI promotes a strategy of defense when it comes to communicating with exponents of digital hate culture, as enshrined in the first four points of the group's codex:

> "I. Human dignity is inviolable. […]
>
> II. We are not against 'them'. We want to solve problems together and act guided by mutual respect, love and reason.
>
> III. We do not wage war but seek conversation.
>
> IV. We never become personal, respect the privacy and personal information of all and waive assumptions." (Reconquista Internet, 2019)

RI rejects the authoritative instruction of discussion opponents about what is right and what is wrong. Instead, the group pursues the tactic to look for common grounds on which both discussion participants can agree and build their conversation on. According to RI's instructions, counter speakers should strive to *understand* their opponents by means of active listening, switching perspectives, and always interpreting the opponent's arguments in the most favorable way possible. RI's communication strategy further stipulates to ask open questions, prevent thematic changes, motivate own opinions, stick to objective criticism, and stay calm and fair. The group emphasizes the significance of always posting responsibly, meaning that users should research at least two reliable sources for every piece of information that they share and, furthermore, that content should always be inspected for potential manipulation and its original source (Ibid.).

Although RG and RI outline how to communicate with the respective counterparty, both groups caution against getting stuck with the deadlocked positions and non-negotiable viewpoints of the respective opponent and, instead, advise to concentrate on the audiences beyond the two opposing camps: "Note: In discussions online, you don't want to convince your enemy, they are mostly pigheaded idiots anyway. The audience matters. And it is not about being right, but being considered right by the audience" (D-Generation, 2018). To win over the audiences, RG employs the tactics of Schopenhauer's *Eristic Dialectic: The Art of Winning an Argument* (Ibid.), which presents 38 ways of being right, regardless where the truth lies. Moreover, the *Handbook for Media Guerillas* advises to be friendly, funny, and to extend a "helping hand" to those audiences who are willing to learn about the "truth". The use of hashtags, such as #wakeup, is suggested to specifically target those wavering between different opinions (Ibid.)

Showing a similar reluctance to waste effort on acting within the "nests of right-wing hardliners" (Reconquista Internet, 2019), RI postulates to concentrate on pages that are visited by the broad majority of users, such as "news outlets, pages with broad reach, pages of public figures, etc." (Ibid.). Being aware of the manipulative communication practices of digital hate culture, RI's strategy mainly rests on uncovering the often-veiled tactics described above so that other discussion participants, as well as silent audiences, realize the tricks that are commonly used to tamper with the art of argumentation. Their strategic communication documents comprise an extensive wiki on the foundations of argumentation. Chapters, for instance, illustrate typical arguments of the far-right and

advise how to confront the argumentation strategies deriving from Schopenhauer's *Eristic Dialectic*. RI does not only promote a thorough understanding of the opponent's strategies but also stresses the role contextual factors play in communication. There are chapters critically reflecting on the significance of own biases and heuristics, communication theory, the basics of Formal Logic, and statistics (Ibid.).

Two general strategies that both groups have in common is that they utilize humor and operate in teams. Postulating that "a laughing enemy is already halfway on our side" (D-Generation, 2018), the *Handbook for Media Guerillas* calls on its readers to be funny and creative in their communication. Two concrete tactics include the creation of satire accounts and memetic warfare. Latter is considered an appealing and hence effective way of "bringing narratives to the people" (Ibid.). Memes are utilized to expose the "lies" of the establishment and to humiliate enemies by, for instance, photoshopping the opponents' profile pictures or mocking them with comic characters, such as Pepe the Frog. That RG appropriates humor and the 'logic of lulz' to veil its actions as being 'merely for fun' becomes clear in the handbook's opening sentence: "We all like to take the piss out of victims on the Internet. […] Here comes a small manual, without any claim to completeness" (Ibid.). Branding the handbook as a manual for "trolling" and "mockery" trivializes the subsequent calls for abusive language and the advocated systematic manipulation of public opinion. It further stands in contrast to the closing words of the handbook that take on a more serious tone when mobilizing to "give the globalists the final push" so that the future may be "reconquered" (Ibid.). When looking at the role humor plays in the communication strategies of RI, instructions pose a clear set of rules on what can and cannot be joked about. RI states that memes should never contain swear words, insults, defamations, or violent references and must never target single individuals (Reconquista Internet, 2019). Regarding the use of hashtags, RI's strategy is to avoid using formulations that point to their identity as an organized counter speech group, since an accumulation of such hashtags would provide a pattern that is easily recognizable for page hosts and algorithms which might lead to blocking (Ibid.).

Working in a team, for RI, means to "hold together, help one another and not allow [themselves] to be divided" (Ibid.), as the codex clarifies. The use of fake accounts is recommended for self-care reasons and to avoid doxing. By contrast, teamwork in the case of RG is promoted as a means to take advantage of algorithmic processes and push

content more efficiently. The *Handbook of Media Guerillas* recommends to set up at least two to three inconspicuous profiles to be able to perform with several accounts simultaneously and infiltrate "the filter bubbles of the average citizen" (D-Generation, 2018). With certain hashtags, fellow members are supposed to be mobilized on internal communication channels like Telegram or Gab to join the "information wars" fought on mainstream social media. The handbook further promotes a three-step model of creating a "swarm intelligence," intended for those who "feel intellectually capable of doing so" (Ibid.). A swarm of users should first extensively share information on a chosen topic by reposting articles, then simplify the information for audiences and, lastly, research additional information and publish such in own articles that are shared across the web (Ibid.). The encouragement of such practices points to the danger of *ungovernability* of digital hate culture that results from its decentralized swarm structure, its ability to quickly migrate across the web, and its use of coded language (Ganesh, 2018).

With regard to the overall research objective of this thesis, the analysis of the strategic communication documents of RG and RI discloses several crucial points. Firstly, it reveals that the advocated counter practices of RI work towards a reduction of exclusionary antagonism by meeting opponents in a respectful, open way and by rejecting moral assessments of 'right' and 'wrong' and polarizing delimitations of a 'we' against 'them'. The practices called for within the realms of digital hate culture do, however, the exact opposite: RG understands its opponents as enemies, not adversaries, without legitimate political standpoint and hence mobilizes its proponents to fight them with all means, ruling out any chance of having a discussion climate necessary for productive agonism. Furthermore, the analysis draws attention to the fact that both organized groups do not actually aim for resolving conflicts that emerge between users spreading hate and those countering it; instead, the communication practices of both sides are mainly striving to affect those users who enter or observe the conflict without positioning themselves on one of the 'extreme' sides. In doing so, RI relies on knowledge-sharing and objective, factual reasoning, while RG focuses on obfuscation, argumentative manipulation, and emotionalizing content. Having analyzed the officially promoted strategies of these two organized groups, the analysis now moves from clear-cut intentions to the more complex mesh of practices actualized on Facebook.

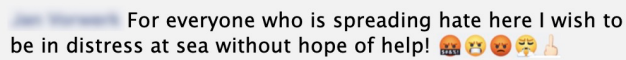### 7.1.2 Actualized communication in online discussions

The aspired communication practices of RG and RI do not equal the totality of practices played out in the comment threads. Certain strategies and tactics described above can surely be recognized, indicating the presence of users who identify with either the culture of digital hate or counter initiatives. However, there are only a few cases in which a clear allocation to organized groups can be made, the most straightforward example being users tagging their posts with hashtags, such as #iamhere which stands for one of the biggest counter speech movements in Europe. The presence of hashtags commonly used in the circles of digital hate culture, such as #weareevenmore or #idrinkbeer[14], allow similar assumptions to be made for the opposing side. Looking beyond these cases, however, the majority of comments does not allow an unambiguous division in 'organized hate group' and 'organized counter initiative'. Instead, the analysis of user comments on Facebook illustrates that hate speech and counter speech can take different forms – some of which appear more efficient for reaching the respective goals than others.

While digital hate culture relies on obfuscating their manipulative practices making it hard, if not impossible, to tell if someone is following actual strategies of an organized network or is simply spreading prejudices and insults on his or her own agenda, counter initiatives, such as RI, communicate their strategies and tactics openly making it obvious when counter efforts deviate from the advocated practices. Thus, the counter efforts present in the analyzed user discussions must be distinguished in counter speech, as promoted by RI and recognized by scholars as respectful, true, fact-bound, and thoughtful reasoning (e.g., Schieb & Preuss, 2018) on the one hand, and such counter efforts that challenge hate speech but do not follow the strategies described in the previous chapter. The latter form of counter efforts dominates the discussions analyzed. Strongly deviating from the instructions postulated by RI, counter speakers, for instance, often act in an authoritative manner, instructing their discussion opponents about what is 'right' and what is 'wrong' and quickly assess users personally as being stupid, lacking morals, etc. Such behavior only leads to a hardening of standpoints instead of fostering constructive exchange, as RI warns, and confirms the negative potential of counter efforts to further pave the way for hate speech (Ernst et al., 2017). This goes so far that some instances of counter efforts showcase similar characteristics as the practices promoted by digital hate

---

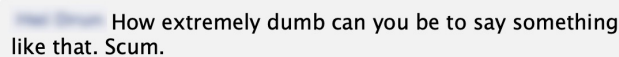[14] Their meaning will be elucidated at a later stage.

culture (Figure 6). This is, for example, evident in the use of expressive speech acts that Jaki and De Smedt (2018) found to be typical for hate speech.
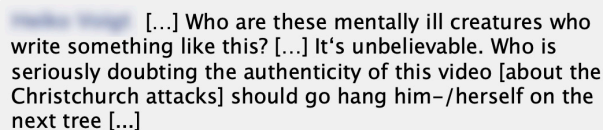
> For everyone who is spreading hate here I wish to be in distress at sea without hope of help! 😱😡😠🤮🙏

Figure 6: User comment example I

Instead of furthering a constructive exchange of arguments, such speech acts provoke the discussion opponent and are emotional rather than supporting an objective and factual exchange. In general, much of the counter efforts present in the discussions go against the basic standards that RI enshrines in their codex: counter speakers target opponents personally, make provocative, emotional, and pedantic comments, and are not inhibited from using abusive language (Figure 7).

> How extremely dumb can you be to say something like that. Scum.

> […] Who are these mentally ill creatures who write something like this? […] It's unbelievable. Who is seriously doubting the authenticity of this video [about the Christchurch attacks] should go hang him-/herself on the next tree […]

Figure 7: User comments example II

To describe and differentiate between the different roles users take in the analyzed discussions, it is helpful to draw on the distinction Schieb and Preuss (2018) suggest. The *core* – users who propagate extreme opinions and cannot be influenced by counter speech (Ibid.) – and the *counter speakers* (in their different forms) are the most dominant participating user groups. The *followers* – the undecided audience who is easier to be influenced by counter speech (Ibid.) – is least active. This means that the vast majority of the participating users has a firm opinion about the topics discussed. There are no instances of comments indicating that a user is uncertain, looks to form an opinion, or shows a willingness to change her or his own opinion. Instead, the analysis suggests that most users who decide to engage in a Facebook discussion like the ones analyzed, identify with one of the opposing positions and get active to defend their position rather than to seek exchange. This raises interesting implications in connection to the objective of both

organized counter initiatives and groups of digital hate culture to primarily address the undecided audience and avoid pointless confrontations with the counterparty. If the susceptible user that both sides wish to convince is not active in the discussions, the efforts made consequently come down to counter-arguing the opponent's statements in a way so that the silent audiences – the readers of the discussions – may be convinced of the truthfulness and validity of one of the opposing standpoints. However, this dynamic only intensifies the need to 'defeat' the discussion opponent by all means, which reinforces the polarization of the competing standpoints and does not benefit an agonistic discussion environment as the formation of discourses plainly shows.

7.2 Discourses

7.2.1 The struggle over the 'truth' about immigration and Islam

Essentially, the discussions between the two opposing sides evolve around different representations of the 'truth'. Just as the analysis of strategic communication documents suggested, it appears to be the most central objective of the users to defend the 'right' representation of the discussion subject to the reading audiences. While one side promotes discourses of hostility and rejection towards Islam and Muslim refugees, the other side calls on human values and promotes equality, tolerance, and openness.

The range of negative representations of refugees and Islam found in the discussion threads illustrate how cyberhate can generalize and denigrate an entire group of people, using "threats, discrimination, intimidation, marginalizing, otherings and dehumanising narratives" (Blaya, 2018, p. 2). The Islamophobic statements thereby resemble the themes identified in anti-Muslim hate speech: the rejection of refugees, concerns about a cultural displacement due to an 'Islamic invasion,' the threat of Islamic terrorism, and a general Islamophobic world-view (Davey and Ebner, 2017). Recurring allegations maintain that Muslim refugees and immigrants are violent, criminal, backward, intolerant, uneducated, culturally strange, and only seek to exploit the European welfare systems. The targeted group of people is designated as rapists, misogynists, fundamentalists, illegals, invaders, and as a general threat to Western culture, illustrating the high level of abusive language in anti-Muslim hate speech (Jaki and De Smedt, 2018).

The discourse that is perpetuated through such verbal abuse, despise, prejudices, and rhetoric based on false assumptions, antagonizes refugees and Muslim immigrants, building up the image of an enemy that needs to be fought against. Hate speech thereby strongly contributes to the process of 'othering,' creating a binary opposition between Islam and Christianity which suggests the incompatibility of Muslims with Western culture. Representations that enforce this discourse work through the different communication practices of digital hate culture outlined above, such as the persuasion techniques of the *Eristic Dialectic*. In the following example (Figure 8), a statement on the high number of refugees who died at sea is being relativized by misleading comparisons and so-called 'whataboutism,' through which topics are switched by referring to a seemingly connected but for the original statement irrelevant topic (Reconquista Internet, 2019).

Figure 8: User conversation example I

The use of misinformation supports the anti-Muslim discourse further. Conspiracy theories are often interwoven with one another in the attempt to make a sound argument for the reader, relating to what Ganesh's (2018) line of reasoning that digital hate culture builds "on a cultivation of common sense amongst its audiences that ultimately seeks to radicalize those who listen" (p. 33). The discussion under the article about sea rescue missions for refugees is pervaded by comments suggesting that the majority of immigrants who came to Europe since 2015 does not classify as actual 'refugee'. A popular explanation for people's 'true' motivation to leave their home is the appeal of the social welfare systems of European countries. This allegation is underpinned by the

conspiracy theory that Europe is, in fact, luring people to come here as governments financially benefit from the booming asylum industry (Figure 9).

> This has nothing to do with refugees. They don't want to integrate, they only want in the social hammock.

> [...] and who or what do they allegedly flee from? They are being attracked by you. The word has spread to every last corner of Africa that there are boats of social justice warriors waiting in the Mediterranean playing water taxi to Europe, the promised land.
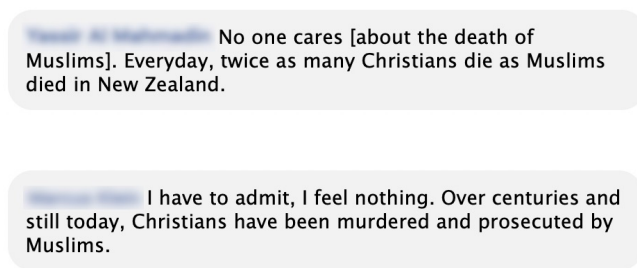
Figure 9: User comments example III

The overarching themes of such conspiracy theories express anxieties about losing control within a political, religious, or social order (Marwick & Lewis, 2017). What becomes obvious in the allegations made throughout all analyzed discussions, is the strong "distrust of government or the 'official stories' of the media […] driven by a belief in the machinations of a powerful group of people who have managed to conceal their role in an event or situation" (Ibid., p. 18). This constitutes an anti-establishment discourse that, to Mouffe (2005), is directly connected to the dangers of the consensus model of the prevalent post-political *Zeitgeist*: "the weakened left/right opposition [is replaced] by a new type of we/they constructed around an opposition between 'the people' and 'the establishment'" (p. 70). The anti-establishment discourse is reflected in several of such conspiracy theories throughout the analyzed discussions: there is the idea that Europe is plotting and financially benefiting from the influx of refugees, the allegation that the reported survey about Germans' attitudes towards immigration is faked by the media and the responsible research institute, and the hypothesis that the Christchurch attacks have been plotted by the 'system' to make potential future attacks by Islamic fundamentalists appear less severe. These narratives relate to the 'White genocide' and 'Red Pill'[15] tropes that Ganesh (2018) explains to be essential for connecting the fractured groups of digital hate culture through a 'common spirit'. To

---

[15] The 'White genocide' conspiracy theory contends that "Western civilization and culture is facing an existential threat" (Ganesh, 2018, p. 34) from mass immigration and racial integration that is believed to be a deliberate action aiming to dismantle white collective power. The term 'Red Pill' is used as a metaphor derived from the movie *The Matrix* where taking the blue pill means to live a life of delusion and taking the red pill means to become enlightened to life's ugly truths. Being 'red-pilled,' in the context of digital hate culture, means to realize the deceit committed by the 'leftist project' (referring to feminists, Marxists, socialists, and liberals) who conspire to destroy Western civilization and culture (Ibid.).

unknowing audiences, such 'insights' appear to reveal a hidden truth that the establishment is withholding, namely, that mass immigration is an 'existential threat' to Western civilization, making it a powerful tactic to convert others to totalizing and extreme worldviews. Counter voices trying to oppose such misinformation fit well into the narrative: any efforts to brand exponents of cyberhate guilty of hate speech or censor their content "only strengthen their resolve and buttress the claim that they speak a truth that is being suppressed by power" (Mouffe, 2005, p. 37).

When accepting any of such presuppositions that suggest the betrayal of the people by the establishment, it becomes easier for users to incite hatred and somewhat more acceptable to not show empathy for the hardships that 'others' go through. In case of the Christchurch attacks, it appears to be sufficient for users to refer to violent acts carried out by Muslim perpetrators in the past to justify a lack of empathy for the 49 people who died in the attack (Figure 10).

Figure 10: User comments example IV

Such simplistic depictions of alleged correlations show, again, how digital hate culture works to defend its positions towards the reading audiences. When presented in form of memes, false allegations and conspiracy theories can be communicated in an even more simplified and emotionally attaching way, making it a powerful tool for exponents of digital hate culture to grab the readers' attention. Here is an example for such a meme that falls into both anti-refugee and anti-establishment discourses, maintaining that if people have enough money to pay the tug boat and if the majority of them is men but not women or children, then their plight cannot be as serious as the media is telling people (Figure 11). This example highlights the dangerous potential of memes that, when used strategically, they are often conceptualized in a way that they appeal to "multiple

audiences far beyond those who unambiguously identify with neo-Nazi and other far-right symbolism" (Bogerts & Fielitz, 2019, p. 150).

Figure 11: Meme example I

"As long as you bring those, into our country instead of those, you hypocrites can kiss my ass!"

The dichotomizing discourse of juxtaposing Islam against the 'West' or Christianity goes hand in hand with a strong moralization of the two sides, framing them as 'evil' and 'good'. How such moralization can become the essence of a discussion becomes most obvious in the comment thread under the article about the Christchurch attacks. The moralizing discourse emerges from representations that generalize groups of people based on their religious identity, emphasizing how the respective religion has endorsed violence and terror in the past. Such representations allow users to relativize the recent deaths of Muslims by listing earlier terror attacks through which Christians have died. This evolves into some sort of 'competition,' in which users from each side try to outdo each other with numbers of victims that can be ascribed to the wrong-doing of either Islam or Christianity (Figure 12).

A small selection:

January 2019, Philippines: 27 murdered Christians in a cathedral.

2017/18, Egypt: 35 murdered Christians on their way to the church.

December 2017, Pakistan: 40 dead after an attack on a Christian church.

June 2014, Kenia: about 50 murdered Christians etc.

Any special broadcast in Germany? Nope.

Condemnation by the chancellor and other members of the government: nope.

But now A. Nahles [Leader of the Social Democratic Party] is immediately declaring: "Attacking Muslims (on the other side of the world) means attacking us!"

[…] Afghanistan 14 years of war: 220.000 dead

Irak 12 years of war: over 1.000.000 dead

Lybia 4 years of war: 40.000 dead

Syria about 4 years of war: 250.000 dead

These are the official numbers, who knows how high the real figures are.

[…] 2,5 Million murdered Christians between 1894–1920 in Turkey, Greeks, Arameans, Armenians; 30.000 murdered Dersim-Kurds, 500 destroyed churches in Northern Cyprus. Any further questions???

Figure 12: User conversation example II

These lines of argumentation are clear instances of what Darmstadt et al. (2019) refer to as toxic narratives. The Christchurch attacks are outweighed by terrorist attacks that have been carried out by Islamic fundamentalists around the world, establishing an insider reference that similar attacks by Christian extremists can be seen as a reaction or revenge and should hence not be condemned in the same way. Narratives, as Darmstadt et al. explain, can provide necessary frames for interpretation and help to establish connections between isolated events but they also "stir up emotions and can help to motivate and mobilize" (Ibid., p. 160), which makes them valuable tools for sowing hatred and fear. By recurrently postulating correlations and causalities, exponents of digital hate culture trigger such emotions and make narratives toxic to society (Ibid.). Such toxic narratives, then again, contribute to transforming hate speech into a form of dangerous speech, that increases "the risk that audiences will condone or participate in violence against the targeted group" (Albrecht, et al., 2019, p. 8). Expressions that justify the murdering of Muslims as a form of revenge highlight the violent potential of hate speech. The Christchurch attacks themselves are a clear example that online hate speech and offline physical violence are linked.[16]
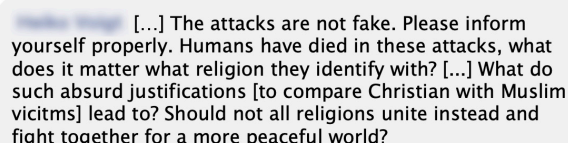
The narratives entail a strong polarization of the two religions and endorse an antagonism that appears to be far from a productive agonistic discussion. The moral framing of the conflict illustrates Mouffe's (2005) thesis that "nowadays the political is played out in the

---

[16] Shortly before the shooting, the attacker and right-wing extremist Brenton Tarrant shared a document on the message board 8Chan, in which he detailed his anti-Islamic and anti-immigration reasons for the attack, making references to typical *alt-right* narratives and the associated Internet culture. Members of 8chan later celebrated the attacks and called for more violence for the white nationalist cause (Wendling, 2019).

moral register" (p. 5), establishing we/they discriminations in terms of 'right' and 'wrong'. The identification with an imagined community of 'us' and 'them' contributes to conveying the impression that neither side sees the standpoint of the opponent as a legitimate one, ruling out a discussion climate of agonistic pluralism. These cases of anti-Muslim discourse point to the limitations that the concept of agonistic pluralism meets when applied to hate speech: as both Cammaerts (2009) and Davids (2018) have emphasized, one needs to consider the extent to which tolerance can offer a plausible response to hate speech. Hate speech, as exemplified above, "intends to be hateful and hence, harmful" (Davids, 2018, p. 306). Davids rightly argues that it only serves "those, who wish to assert their power, and maintain their dominance through the subjugation and deprecation of others. As such, hate speech stands in contradistinction to the values of a democratic framework of regard and respect for human dignity" (Ibid., p. 306). By building their arguments on prejudices and depreciations of certain groups of people, it is hard, if not impossible for counter speakers to find common grounds based on which a productive discussion with mutual respect can build upon. Nevertheless, there are counter efforts present in the analyzed discussion which strive to reduce the antagonism prevalent in the discussions. Challenging the legitimacy of the dominant anti-Muslim and anti-establishment discourses, these efforts can be understood as counter-discourses (Neumayer & Valtysson, 2013).

By recurrently posting links to Wikipedia where the workings of 'whataboutism' are explained, one counter speaker draws attention to the misleading and manipulative ways in which the discussion on the Christchurch attacks is framed. Others emphasize the uselessness of drawing on comparisons of the number of victims, underlining that it is, after all, *humans* who the opposing sides talk about, making it irrelevant what religion they might identify with. Users also step in to counter-argue conspiracy theories and provide links to information sources (Figure 13).

[...] The attacks are not fake. Please inform yourself properly. Humans have died in these attacks, what does it matter what religion they identify with? [...] What do such absurd justifications [to compare Christian with Muslim vicitms] lead to? Should not all religions unite instead and fight together for a more peaceful world?

[...] Before uttering criticism, it is first important to know the basics about how surveys are conducted. The concept of 'representativity' is crucial to understand.
https://de.wikipedia.org/wiki/Repr%C3%A4sentativit%C3%A4t

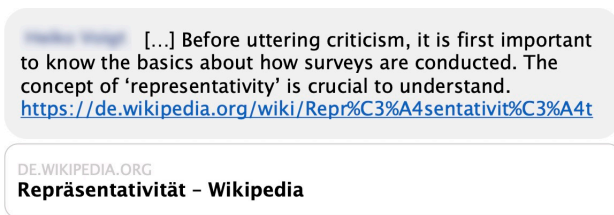DE.WIKIPEDIA.ORG
Repräsentativität – Wikipedia

Figure 13: User comments example V

Although there is a wide range of counter-strategies that can be observed from the discussions, the analysis also underlines their limitations. Calls for a more objective and fact-based discussion and the refutation of misinformation and prejudices does, in the majority of cases, not show any effect on the discussion opponent. Just as Miškolci et al. (2018) outlined, the discussion opponents typically insist to have enough information already and thus see no need to reconsider their opinion. In this regard, it is interesting to look back to Sponholz' (2016) argument who postulates that counter speech merely boosts hatred as it fails to lead to consensus or refutation of hateful contents but instead provides it with legitimacy and relevance and thus supports its discourse quality. Assuming that the Islamophobic and hateful content in the analyzed discussions would just be ignored, abiding by the motto 'do not feed the troll,' and not be met by any form of counter efforts, those audiences who are not aware of the manipulative workings of hate speech would be left to assume that the articulated representations mirror the actual public opinion. Thus, the counter efforts do fulfill an important purpose, even if they are not successful in changing the opponents' opinion. The discussability promoted by organized counter speakers is, moreover, crucial when thinking of Mouffe's (2005) line of argumentation that political positions need a place where they can be expressed and accepted as legitimate standpoints in order to prevent them from developing into extremist positions. Considering these aspects, Sponholz' reasoning appears insufficient to grasp the full potential of counter speech, which, in the light of this analysis, can indeed help to challenge discourses propagated by digital hate culture. Of course, as already stated in the previous chapter, counter speech also carries a risk to further pave the way for the dissemination of hatred. It thus remains a question of *how* counter efforts are carried out in order to assess how it might be able to reduce the antagonism fostered in discourses perpetuated by digital hate culture.

7.2.2 Antagonizing representations of 'left' and 'right'

Next to exploiting the fear towards the unknown cultural and religious 'other,' the practices of digital hate culture in the analyzed discussions also vigorously target the establishment and its 'leftist' proponents, fostering an antagonistic discourse which separates society into a 'left' and 'right' block that are mutually incompatible. In the comments addressing the opposing camp, the strategies and tactics promoted in the *Handbook of Media Guerillas* become visible. The essence of trolling, i.e. means of provoking, discrediting, and humiliating the discussion opponent is especially present. The repertoire of insults thrown at the discussion opponents is wide, reaching from typical name-calling used in far-right circles to address the 'left' as, for instance, social justice warriors (*Gutmenschen*), colored-naïve (*Buntnaive*), and leftist (*Linksversiffte*) to more general slurs as, for instance, calling users maniac, parasite, or victim. Occasionally, the language goes beyond insults, threatening the discussion opponent and making violent references (Figure 14).
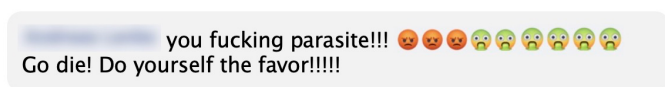
Figure 14: User comment example VI

Some users try to trigger provocation by turning the charges of the opposing side against them, as the *Handbook for Media Guerillas* advises, or by displaying indifference to counter speakers' attempts to disprove hate speech (Figure 15).
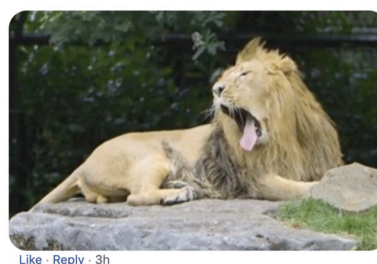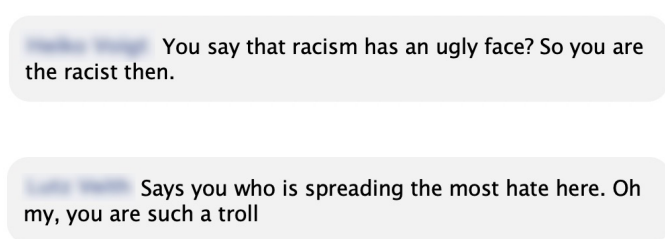
Figure 15: User comments example VII

The use of provocative and abusive language is, however, not reserved for those displaying hostility towards refugees and Muslims. There are many instances to be found in which counter speakers work with insults, harsh language, and humiliation (Figure 16) – clearly showing the negative potential of counter efforts to provide the grounds for

exponents of digital hate culture to turn accusations and enforce antagonisms. Frequent terms used to disclose the ideology of the discussion opponent by counter speakers include, for instance, misanthrope, racist, right-wing extremist, neo-Nazi and 'blue-brown' (referring to the colors associated with the *AfD* and Nazism).

Of course, the notorious right-wing extremist mob is primitively stirring up hatred again.

[…] oh, a sturdy neo-Nazi! Do you actually piss into your sweatpants when you write something like this? 🙍

[…] Detective "hitler salute" has solved the case.
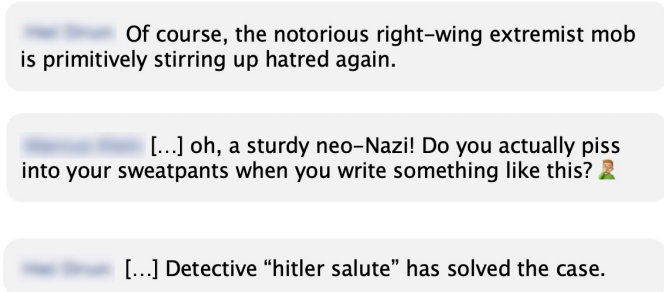
Figure 16: User comments example VIII

Such charges only provoke a back and forth of accusations without any constructive content. This form of counter speech, that does not follow the principles advocated by RI and instead relies on similar provocation practices as digital hate culture, rules out the possibility to accept the discussion opponent as a legitimate adversary. In some cases, memes are appropriated to display the mutual despise in a humoristic way (Figure 17).

"Hate is the orgasm of the non-caressed" (picture: Björn Höcke, AFD politician)

"I found the heart of a AfD supporter"

"#wearemore – stand up against right-wing hatred!"

"Reality denier. #wearemore – stand up against right-wing hatred!"

Figure 17: Memes example II

While the first two memes attempt to counter hate speech by ridiculing the right-wing populist party *AfD*, the other ones are the evoked responses by exponents of digital hate culture. Those memes appropriate the hashtag #wearemore (originally used by counter initiatives to emphasize their majority over users spreading hate) but manipulate the visual content targeting the identity of counter speakers: the first meme aims to ridicule the adherents of the left camp, the second meme contains an edited profile picture of a counter speaker (subsequently blurred) labeled as 'reality denier,' w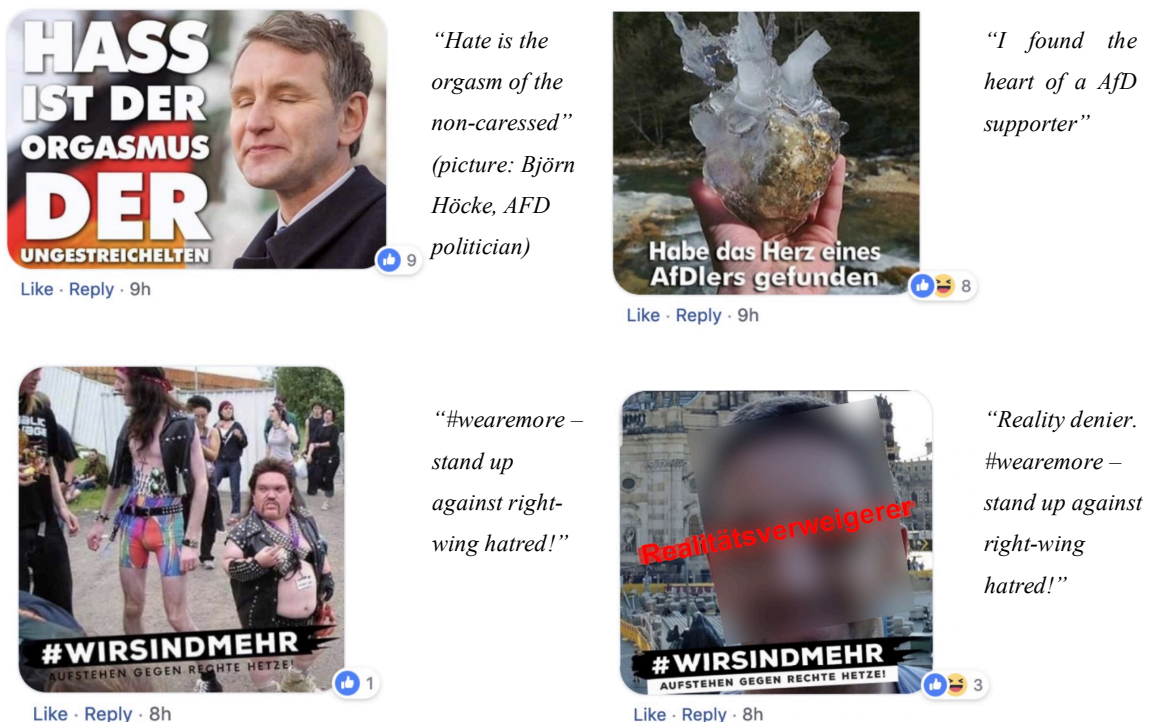hich plays into the 'Red Pill' trope suggesting that the 'leftist project' is denying the people the truth about the threats facing Western civilization. Other memes feeding into the anti-establishment discourse are those declaring information as fake news (Figure 18), a tactic that has been promoted in the *Handbook for Media Guerillas*.

Figure 18: Meme example III

This behavior is a clear incidence of 'trolling,' where users aim to deliberately bait people with inflammatory content in order to elicit an emotional response. The 'logic of lulz,' however, is not only used by the trolls of digital hate culture but also by the counter side. This connects to Milner's (2013) research that ascertained the potential of the 'logic of lulz' to afford adversarial pluralism over exclusionary antagonism. Contrary to Milner's reasoning, the examples found in the analyzed discussions do not indicate that intentional sowing of discord encourages a vibrant agonistic discussion. Rather, as exemplified by the memes shown above, the 'logic of lulz' further antagonizes core identity categories and essentializes binary dimensions of 'left' and 'right'. Although memes – and the 'logic of lulz' that underlies them – provide counter speakers with a powerful tool to appropriate Internet culture in similar ways as digital hate culture does to appeal to audiences (and especially younger generations), it appears nevertheless not to be a productive means to foster true dialogue in an adversarial sense.

The nonconstructive nature of the 'logic of lulz' is also apparent in the appropriation of hashtags by exponents of digital hate culture. Hashtags are repeatedly used to mock and ridicule hashtags originally used by the opponent: #iamhere is transformed into #idrinkbeer and #wearemore is reformulated into #wearemorethandumb, #youareconfused, and #weareevenmore which illustrates the main concern of both groups: to compete for showing the audience who builds the actual majority of the public. The appropriation of hashtags creates in-group jokes, a shared practice that Davey and Ebner (2017) identify to be essential for the convergence of different far-right actors. The plentitude of ironic and sarcastic comments, memes, and mocking appropriations of hashtags highlights again that humor plays a crucial role in the communication of both exponents of digital hate culture and counter speech initiatives.

The outlined practices clearly showcase the 'left and right' distinction present in today's society and refute the aspired post-political vision that seeks to go beyond such antagonisms. Again, what looms in this antagonistic discourse, is the strong moralizing tone that Mouffe (2005) cautions against, similar to the one present in the previously discussed discourse that antagonizes Islam and the 'West'. Both sides seem to seize every opportunity to discredit the opponent by labeling users as either 'part of the deceitful system' or 'right-wing extremist'. This naturally entails distinctions of 'we/them' and 'good/evil'.

"How Germany abolished itself once before: through immigration – through Nationalism"

"If today being a Nazi means to protect one's family, defend one's achievements, make order, fight injustice, challenge arbitrariness, have an opinion, question the political circus, and reject war, then I am happy to be a Nazi!"
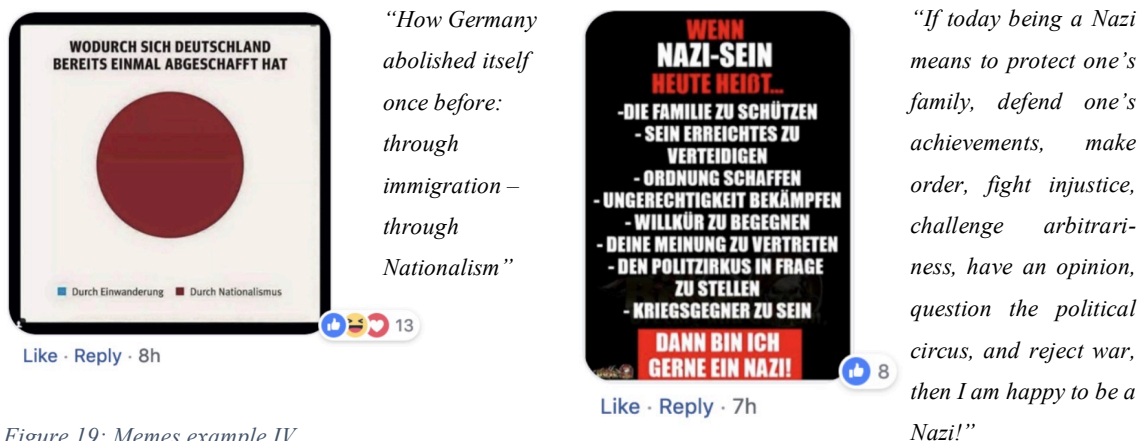
Figure 19: Memes example IV

These two memes (Figure 19) are not reactions to one another within one discussion thread but symbolize how moralizing references to nationalistic sentiment or affinity to right-wing conservative ideology are used by exponents of digital hate culture to benefit their own discourse: the meme to the right implies that it is tolerable to be a Nazi in

today's society, playing with the tendency of counter speakers to use the label thoughtlessly. This is a clear example of how images are employed "strategically in order to disseminate ideology in more or less subtle ways and to persuade others to share or reject certain views and values" (Bogerts & Fielitz, 2019, p. 138). Consequently, memes, despite their humorous, sometimes silly or absurd appearance, are often not as harmless as they appear at first sight. They can convey key ideological narratives, attract new supporters, and contribute to rendering extremist thought as mainstream (Ibid.). Organized counter initiatives, such as RI, caution against practices of 'othering' and framing opponents as a moral enemy using labels that assign political or ideological stances to users. Attributing labels, such as 'Nazi' or 'racist' to users results in simply being dismissed or serves to blame the sender of inciting hatred him-/herself, as the analysis shows. It is, therefore, crucial to raise awareness among counter speakers that taking too extreme positions is counterproductive for the persuasiveness of their cause (Schieb and Preuss, 2018). This, again, underlines the importance to differentiate between organized rule-guided counter speech and general counter efforts.

The analysis of discourses has disclosed a rigor and animosity in which the discursive struggles on Facebook are fought out that leaves no doubts that an understanding of social media as a platform for rational consensus-making in an Habermassian sense is misleading (Cammaerts, 2009). Rather, the clash of opposed positions confirms that antagonisms are constitutive for society: they "can take many forms and it is illusionary to believe that they could ever be eradicated" (Mouffe, 2005, p. 30). Facebook thus functions as a battlefield where rivaling groups carry out a fight over dominating discourses. The constitutive and performative power of such user interactions stress the destructive forces that disenfranchising antagonism unleashes. The analysis has shown that practices of digital hate culture encourage discourses of enmity and antagonistic rivalry in two directions: On the one hand, acts of hate speech promote hostility and fear towards Muslims as cultural and religious 'others,' fostering an understanding of a reality in which people are valued less because of their belonging to imagined communities determined by ethnicity and religion. On the other hand, the provocation and humiliation of political opponents distribute to an anti-establishment discourse that works to delimit the 'deceitful leftist system' from the 'betrayed people'. However, the analysis has also shown that counter-discourses can fuel such polarization of 'good' and 'evil' by stepping into similar moralizing and subjective communication practices.

7.3 Facebook's affordances and power relations

The previous chapter evinced that Facebook has become a key arena for contemporary discursive struggles around identifying the 'true' good and evil of society. It is, therefore, pivotal to consider the affordances and accompanying power relations arising from the platform itself as such can both facilitate and constrain the formation and perpetuation of discourses (Uldam & Kaun, 2019). Facebook constitutes the space in which users defend their political opinion and become part of 'networked publics,' to speak with Neumayer's and Valtysson's (2013) words. Such publics emerge as imagined collectives through networks provided by different media platforms (Ibid.). For the cases analyzed, this means that the imagined collectives of a 'right' networked public spreading hate and propagating against Islam and the establishment on the one side, and a 'left' networked public countering these discourses, are the result of intersections of users, practices, but also, crucially, technology. Constituting the space for the struggle between these networked publics, Facebook facilitates and limits the communicative potentials of the users and their practices through its technical functionalities and affordances.

Some of the practices that organized groups of digital hate culture and counter speakers advocate, heavily rely on Facebook's affordances. The response to comments in form of *reactions* (i.e. the 'like,' 'love,' 'laugh,' 'surprise,' and 'anger' emoji) does not only show a user's opinion on a statement but also works to upvote content, so that certain comments become more visible than others. When scrolling to the comment sections underneath an article, one can, usually by default, only see the 'most relevant' comments of a discussion. Facebook's algorithms consider comments 'most relevant' if a comment was reacted on by friends of the reading user, comes from verified profiles and pages, or, crucially, has many 'likes' and replies (Facebook, 2019a). This impacts discussions in a way that users who are aware of the algorithm's workings can affect or deliberately manipulate which comments will be shown to other users by 'liking' or replying to comments that they support. The comments analyzed for this thesis were chosen by selecting the setting 'all comments,' not 'most relevant,' which is why no direct comparison can be made about how different the first impression on a discussion would be depending on these settings. However, it is important to remember what Schieb and Preuss (2018) pointed out in their study: the first comment has an advantage in moving the average opinion of the

discussion, making this default setting crucial to the interplay of digital hate culture and counter speech.

Both counter initiatives and groups of digital hate culture benefit from algorithmic politics. Comments that are tagged with the counter hashtags #iamhere or #wearemore result in significantly higher supportive reactions ('likes' and 'love') and often display responses of agreement, indicating that counter initiatives are aware of the affordance and work towards an upvoting of fellow counter speakers' content (Figure 20).
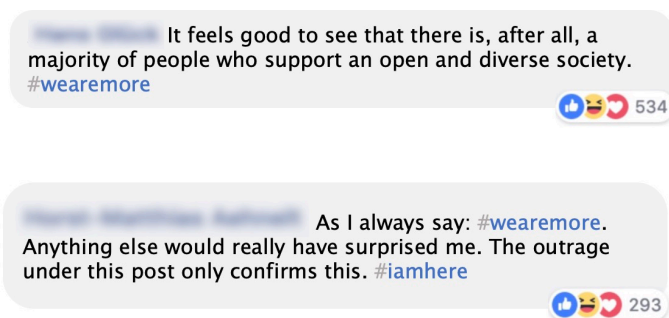

Figure 20: User comments example IX

Looking at the opposing side, the *Handbook for Media Guerillas* clearly displays how digital hate culture relies on the exploitation of platforms' algorithmic politics. The practicability of the advocated 'social networking raids' and 'memetic warfare' depends on a thorough knowledge of algorithmic workings: "In Germany, you sometimes only need 1000 tweets per hour to make a hashtag trending on Twitter (in the US it's 20,000/h). […] If 20 people join, everyone has to post about 50 tweets per hour (less than one per minute). That's possible" (D-Generation, 2018). Research has confirmed that the execution of such coordinated attacks helps organized groups of digital hate culture to reach the top trends of social media platforms and, in this way, dominate online discourse (Kreißel, et al., 2018). From such research it can, moreover, be assumed that 5 % of all accounts actively spreading hate speech on Facebook generate 50 % of the likes for degrading comments (Ibid.). Here, another platform affordance benefitting the cause of digital hate culture comes into play: the ease of using fake accounts and the possibility to deploy social bots. Latter can be understood as "pieces of software that create content on social media and interact with people" (Marwick & Lewis, 2017, p. 38). Social bots are increasingly used for media manipulation, as they allow to inflate the numbers of followers or reactions and help to aggregate and spread propaganda. They are cheap and

easy to deploy and "it is often difficult for average users to distinguish between 'real' users and bots" (Ibid.), making them a dangerous tool that benefits the proliferation of cyberhate. These affordances combined allow a loud minority to suggest a false majority opinion on mainstream social media, which connects to Matamoros-Fernández' (2017) observations and shows that Facebook can indeed be understood as a space facilitating 'platformed hatred'.

Users who are aware of the algorithmic politics of Facebook, consequently have a clear advantage as they can appropriate the affordances to push their own agenda. The same applies to the platform policies that stipulate what can and what cannot be said. The strategic communication documents of RG confirm that exponents of digital hate culture are well aware of social media platforms' policies and legal boundaries that need to be circumvented to avoid blocking: "Don't make any criminally relevant statements […]. Don't threaten to use violence, but make your opponent do it. Then you can report him/her and let him/her be blocked" (D-Generation, 2018). Facebook's community standards prohibit hate speech and feature a policy rationale listing in detail what users are not allowed to post, including some of the content that was identified in the analyzed comment sections: "Dehumanizing speech such as reference or comparison to […] [s]exual predator, [s]ubhumanity, [v]iolent and sexual criminals" (Facebook, 2019b). While it is, for instance, also prohibited to maintain multiple accounts and to create inauthentic or fake profiles (Ibid.), the analysis suggests that it is easy to circumvent such regulations. As Schieb and Preuss (2018) explain, there are gaps through which the policies on hate speech can be bypassed: "Facebook's community standards state that hate speech is not prohibited per se, but is allowed under certain circumstances, such as expression of humor/satire, raising awareness for certain topics etc." (p. 283). What is considered 'humor' is not further explained by Facebook. The protection of humor as a guarantor of freedom of expression is problematic since the preservation of ambiguity is a crucial property of current practices of right-wing media manipulation, also referred to as 'Poe's Law'. Poe's Law stipulates that it is difficult or impossible to differentiate between an expression of sincere extremism and a parody of extremism without a clear indication of the author's intent (Marwick & Lewis, 2017, p. 5). Such build-in defenses for satire and irony on social media platforms hence facilitate the spread of toxic narratives.

At large, the benefits that come with knowledge about platform affordances and policies illustrates the power hierarchies involved in the discursive struggles taking place on social media: users who have an understanding of why certain content is privileged over other and are aware that not every commenting user always represents a human individual, have an advantage compared to users who lack this kind of digital media literacy. This emphasizes the need to a wide and large educational campaign that scholars, as well as counter initiatives, frequently call for. What adds up to the power hierarchies at work on social media platforms, is the fact that private corporations carry the responsibility for policy-making as well as monitoring compliance with directives and laws. In this case, this means that Facebook ultimately holds the power over content which is critical for several reasons. Computationally, it is not (yet) possible for Facebook to promptly delete hate speech posts as high accuracy fully automated "hate speech recognition, especially its separation from humorous posts or discussions about hate speech is currently almost intractable" (Schieb & Preuss, 2016, p. 3). This means that it comes down to Facebook's employees to differentiate hateful content punishable by law from acceptable hateful content, highlighting the predicament: successful decoding of ambiguous statements is difficult, especially in online environments, as it "depends on underlying knowledge of the context, the intention, or the social background" (Ibid.) and the daily myriad of reported postings constitutes "too great a burden to the community operations team" (Ibid.). Schieb and Preuss (2016) embark on a thought experiment showing that it is illusionary to rely on social media platforms to regulate hate speech efficiently:

> "let us assume that a native speaker needs about 1 minute to check if a complaint justifies the deletion of a post […], and to actively perform the deletion and/or block the responsible user […] Facebook would need around 100 native German speakers in their community operations teams in order to cope with this amount of complaints." (p. 2)

That regulatory interventions are needed and also expected by users, is reflected in the analyzed material, where users repeatedly call on the page hosts to moderate the discussions and delete hateful content (Figure 21).

TAG24 Dresden – please moderate!

Here? This brings them way more comments than they had in the last months. You don't seriously believe that they would intervene?

Figure 21: User conversation example III

While some page hosts react to these calls or clarify rules from the beginning, such measures cannot be regarded as a comprehensive solution, as it remains an individual decision of the respective page host to interfere that by far not all news outlets follow. As private corporations, most page hosts also pursue their own interests, for some, this might imply that heated discussions are regarded as favorable as they increase the number of active users on their page, as it was pointed out by a user in the example above.

This last analytical dimension sheds light on several points. Firstly, it is important to consider affordances and power hierarchies as technologies must be understood as inherently political spaces "designed for a specific purpose that fosters certain appropriations more than others" (Neumayer & Valtysson, 2013, p.4), thereby privileging those who know their rules. This makes it a welcoming space for both users wishing to spread hatred with manipulative means and for users who try to balance out the formation of false majorities. Secondly, the interplay between these networked publics of digital hate and counter speech is pervaded by power structures that determine what can be said, how, and on which pages. The rules for this are not established by international law or democratic institutions but by private corporations, which fails to comply with Mouffe's (2005) call for democratic politics to provide an organized space in which the plurality of adversarial positions can find expression (p. 20). The fact that the platform affordances essentially support users wanting to hijack social media sites' technical infrastructure for their benefit, as highlighted by Matamoros-Fernández (2017), and thus foster manipulation and privilege, further reasserts that Facebook cannot provide conflicts with the legitimate form of expression needed for establishing productive agonistic discussion.

8 Discussion and concluding remarks

Having looked at the interplay between digital hate culture and counter speech within the four dimensions of practices, discourses, affordances, and power on a particular case study, now allows for reflections on the potential of such confrontations on social media

to facilitate productive agonism over exclusionary antagonism. Above all, the analysis has shown that Mouffe's critique of the post-political *Zeitgeist* is highly justified. The optimistic view that advocates of a consensual form of democracy hold seems illusionary: the examined cases showcase that there is an antagonistic dimension constitutive of *the political* which cannot be negated. The partisan battle fought on social media strongly resists the understanding of a world 'beyond left and right,' as claimed by liberal rationalists. On the contrary, the antagonism highlighted throughout the last chapters accentuates the moment of crisis that democratic politics currently goes through: there is no real partisan debate possible in democratic institutions as parties increasingly seek consensus in the center, not allowing voters "to make a real choice between significantly different policies" (Mouffe, 2005, p. 66). Consequently, right-wing demagogues gain popularity and publics seek other spheres of political expression and debate: the Internet becomes a contested terrain used by 'left' and 'right' to promote their own agendas and interests (Cammaerts, 2009). But can Facebook create "a vibrant 'agonistic' public sphere of contestation where different hegemonic political projects can be confronted" (Mouffe, 2005, p. 3), something that should be the task of democratic politics? The analysis suggests that this is not the case. The dominant practices and discourses display a clash of two opposing groups in which "the we/they confrontation is visualized as a moral one between good and evil, [and in which] the opponent can be perceived only as enemy to be destroyed and this is not conducive to an agonistic treatment" (Ibid., p. 5). The mostly provocative and often offensive contributions do not leave room for productive exchange, opinion building, and mutual orientation. Users are more concerned with manifesting their own position and distinguishing it from the position of others. Contributions of discussion participants are mostly rejected, frowned upon, and ridiculed, creating a discussion climate that only contributes to a further polarization of opinions. This applies both to digital hate culture and such counter efforts that do not comply with the instructions of organized counter speech. Although Facebook, contrary to the channels of deliberate democracy, provides a platform where partisan conflict can find expression and has shown recent efforts to extend its existing bans on hate speech (now prohibiting white nationalist and white separatist content), its technological affordances nevertheless serve the exclusionary antagonism that evolves in it. Without general internationally applicable guidelines and means to handle the manipulative practices of digital hate culture efficiently, Facebook is not a suitable platform to let conflicts emerge in an agonistic form. As private corporations with own interests and no legally regulated

obligations to take action against cyberhate, social media platforms do not offer a promising agonistic perspective.

Nevertheless, Facebook and other social media platforms will continue to be the channels where partisan conflict finds expression whilst democratic politics need to move away from trying to overcome antagonism – as it is constitutive for society and will always exist among collective identities – and instead construct we/they discriminations in a way that passions are mobilized through democratic institutions (Mouffe, 2005, p. 70). Accepting the current condition and its disadvantages, there is still reason to argue that the antagonistic debates can be moved towards productive agonism. The practices of organized counter speech, as advocated by groups, such as RI, and as occasionally found in the analyzed discussions, carry the potential to foster agonistic pluralism in Mouffe's sense: with its thorough knowledge about the opponent's manipulative practices and its appropriation of Internet culture, it provides a strong opposition to digital hate culture. One that is political, not moral, as it advocates to see the opponent as an adversary who has a legitimate political standpoint and seeks mutual understanding. Here, however, the question put forth by Cammaerts (2009) and Davids (2018) remains as to where the limits of tolerant agonism should be when it comes to hate speech. Since a productive agonistic discussion requires a minimal consensus on normative issues, including certain human values that should not be up for debate, the often-dehumanizing practices of digital hate culture certainly exceed these limits in many cases. Following Mouffe's argument, the exclusion of voices from public discourse and labeling them as 'extreme-right,' would only exacerbate the problem. Until there are efficient computational means of automatically detecting those cases of hate speech that clearly disregard and violate basic human values, organized civil-led counter initiatives thus constitute a valuable and necessary means to combat cyberhate. Not only because their practices strive to correspond to the requirements of productive agonism, but also because their efforts speak to the publics beyond those actively participating in online debates: well aware of what crucial role the perception of online debates plays for public discourse, counter initiatives spread an important message, namely that hate does not dominate public opinion and that it no longer goes unchallenged.

The insights gained from this thesis indicate several directions for future research that strives to better understand and undermine digital hate culture. Firstly, the workings of

cyberhate on mainstream social media platforms should be investigated further. Although the hubs for the formation of toxic discourses lay elsewhere on Alt-tech platforms, blogs, chat forums, and in the corners of the dark web, it is on the mainstream online platforms where filter bubbles burst open and exponents of digital hate culture work to manipulate and attract the 'average' citizen. Looking at user comments as a form of political online participation as well as the potential effects of user comments on how people perceive the related online content and expressed opinions is a fruitful path for conducting such studies. An perhaps even more insightful approach, however, would be to investigate this phenomenon from an audience perspective. To better understand what counter efforts are able to unfold and how critically readers of such comment threads reflect on manipulative practices online, it is necessary to include audiences in future research efforts. After all, it is the undecided audience that both digital hate culture and counter initiatives address and aim to convince. Thus, it appears crucial to investigate who the silent readers of discussions taking place on social media are, how they perceive such antagonistic battles between 'good' and 'evil' and what they take away from it.

When it comes to the future of civil-led counter initiatives, this thesis has clearly shown that such organized efforts to combat cyberhate are an effective means that should be widely promoted and supported by more research that confirms its potential and helps to improve its activities. Current recommendations of researchers include that counter speech needs to become even more dynamic (involving both proactive communication and rapid response systems), innovative (advocating out-of-the-box-thinking), and bold (emphasizing the need to break taboos and transcend the limits of conventional debates through humor) (Ebner, 2019). However advanced, counter initiatives cannot remain the only efficient response to the unrestrained hatred and the poisoning of public discourse that diffuses from digital spaces. To be successful in fighting and preventing digital hate culture, "the combined efforts of civil society actors, social media providers, and national legislators are required" (Darmstadt, et al., 2019, p. 164). Building the basis for these response systems, extensive research will be needed, without which it will not be possible to keep up with the rapidly evolving new media ecosystems of the culture of digital hate, their internal dynamics, and their harmful influence on contemporary society.

9   References

Bibliography

Albrecht, S., Fielitz, M. & Thurston, N., 2019. Introduction. In: M. Fielitz & N. Thurston, Eds. *Post-Digital Cultures of the Far Right: Online Actions and Offline Consequences in Europe and the US.* Bielefeld: transcript, pp. 7-22.

Anders, R., 2017. *Lösch Dich! So organisiert ist der Hate im Netz.* [Online] Available at: https://www.youtube.com/watch?v=zvKjfWSPI7s [Viewed 6 April 2019].

Askanius, T., 2019. *Studying the Nordic Resistance Movement: three urgent questions for researchers of contemporary neo-Nazis and their media practices.* [Online] Available at: https://doi.org/10.1177/0163443719831181

Awan, I., 2016. Islamophobia on Social Media: A Qualitative Analysis of the Facebook's Walls of Hate. *International Journal of Cyber Criminology,* 10(1), pp. 1-20.

Awan, I. & Zempi, I., 2017. 'I Will Blow Your Face Off' - Virtual and Physical World Anti-Muslim Hate Crime. *British Journal of Criminology,* 57(2), pp. 362-380.

Ayyadi, K., 2018. *WIE „RECONQUISTA GERMANICA" AUF DISCORD SEINE "TROLL-ARMEE" ORGANISIERT.* [Online] Available at: https://www.belltower.news/wie-reconquista-germanica-auf-discord-seine-troll-armee-organisiert-47020/ [Viewed 6 April 2019].

Blaya, C., 2018. Cyberhate: A review and content analysis of intervention strategies. *Aggression and Violent Behavior,* pp. 1-10.

Bogerts, L. & Fielitz, M., 2019. "Do You Want Meme War?" Understanding the Visual Memes of the German Far Right. In: M. Fielitz & N. Thurston, Eds. *Post-Digital Cultures of the Far Right: Online Actions and Offline Consequences in Europe and the US.* Bielefeld: transcript, pp. 137-153.

Boyd, D. & Crawford, K., 2012. Critical questions for big data. *Information, Communication and Society,* 15(5), pp. 662-679.

Cammaerts, B., 2009. Radical pluralism and free speech in online public spaces: the case of North Belgian extreme right discourses. *International journal of cultural studies,* 12(6), pp. 555-575.

Connolly, W. E., 1991. *Identity/Difference: Democratic Negotiations of Political Paradox.* Ithaca: Cornell University.

Darmstadt, A., Prinz, M. & Saal, O., 2019. The Murder of Keira: Misinformation and Hate Speech as Far-Right Online Strategies. In: M. Fielitz & N. Thurston, Eds. *Post-Digital Cultures of the Far Right: Online Actions and Offline Consequences in Europe and the US.* Bielefeld: transcript, pp. 155-167.

Davey, J. & Ebner, J., 2017. *THE FRINGE INSURGENCY: Connectivity, Convergence and Mainstreaming of the Extreme Right,* London: ISD.

Davids, N., 2018. On the (in)tolerance of hate speech: does it have legitimacy in a democracy?. *Ethics and Education,* 13(3), pp. 296-308.

Deutscher Bundestag, 2019. *Plattform "Reconquista Germanica".* [Online] Available at: https://www.bundestag.de/presse/hib/586296-586296 [Viewed 6 April 2019].

D-Generation, 2017. *DAS SHITPOSTING 1×1.* [Online] Available at: http://d-gen.de/2017/05/shitposting-1x1/ [Viewed 6 April 2019].

Ebner, J., 2019. Counter-Creativity: Innovative Ways to Counter Far-Right Communication Tactics. In: M. Fielitz & N. Thurston, Eds. *Post-Digital Cultures of the Far Right: Online Actions and Offline Consequences in Europe and the US.* Bielefeld: transcript, pp. 169-181.

Ernst, J. et al., 2017. Hate Beneath the Counter Speech? A Qualitative Content Analysis of User Comments on YouTube Related to Counter Speech Videos. *Journal for Deradicalization,* 10, pp. 1-49.

Facebook, 2019a. *Facebook Help Center.* [Online] Available at: https://www.facebook.com/help/539680519386145) [Viewed 15 May 2019].

Facebook, 2019b. *Community Standards.* [Online] Available at: https://www.facebook.com/communitystandards/ [Viewed 15 May 2019].

Ganesh, B., 2018. THE UNGOVERNABILITY OF DIGITAL HATE CULTURE. *Journal of International Affairs,* 71(2), pp. 30-49.

George, C., 2015. Hate speech law and policy. In: R. Mansell & P. H. Ang, Eds. *The International Encyclopedia of Digital Communication and Society.* Malden, MA: Wiley Blackwell, pp. 305-314.

George, C., 2016. Regulating "hate spin": The limits of law in managing religious incitement and offense. *International Journal of Communication,* 10, pp. 2955-2972.

Gollatz, K., Riedl, M. J. & Pohlmann, J., 2018. *Removals of online hate speech in numbers.* [Online] Available at: https://zenodo.org/record/1342325#.XKsOSy-Bqu4 [Viewed 6 April 2019].

Hsieh, H.-F. & Shannon, S. E., 2005. Three Approaches to Qualitative Content Analysis. *QUALITATIVE HEALTH RESEARCH,* 15(9), pp. 1277-1288.

Jaki, S. & De Smedt, T., 2018. *Right-wing German Hate Speech on Twitter: Analysis and Automatic Detection.* [Online] Available at: http://www.organisms.be/downloads/jaki2018.pdf [Viewed 6 April 2019].

Kreißel, P., Ebner, J., Urban, A. & Guhl, J., 2018. *HASS AUF KNOPFDRUCK: Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz,* London: ISD.

Laclau, E. & Mouffe, C., 1985. *Hegemony and Socialist Strategy.* 2nd ed. London, New York: Verso.

Landesanstalt für Medien NRW, 2018. *Ergebnisbericht Hassrede,* Düsseldorf: forsa.

Lauer, S., 2018. *Infokrieg für die Grundschule: „RECONQUISTA GERMANICA" UND „D GENERATION".* [Online] Available at: https://www.belltower.news/infokrieg-fuer-die-grundschule-reconquista-germanica-und-d-generation-46716/ [Viewed 6 April 2019].

Lee, E.-J., 2012. That's Not the Way It Is: How User-Generated Comments on the News Affect Perceived Media Bias. *Journal of Computer-Mediated Communication,* 18(1), pp. 32-45.

Lee, E.-J. & Jang, Y. J., 2010. What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research,* 37(6), pp. 825-846.

Leonhard, L., Rueß, C., Obermaier, M. & Reinemann, C., 2018. Perceiving threat and feeling responsible: How severity of hate speech, number of bystanders, and prior reactions of others affect bystanders' intention to counterargue against hate speech on Facebook. *Studies in Communication and Media,* 7(4), pp. 555-579.

Lincoln, Y. S. & Guba, E. G., 1985. *Naturalistic Inquiry.* Beverly Hills: Sage.

Müller, K. & Schwarz, C., 2017. *Fanning the flames of hate: Social media and hate crime.* [Online] Available at: http://dx.doi.org/10.2139/ssrn.3082972 [Viewed 6 April 2019].

Marwick, A. & Lewis, R., 2017. *Media Manipulation and Disinformation Online,* New York: Data & Society Research Institute.

Massanari, A., 2017. #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society,* 19(3), pp. 329-346.

Matamoros-Fernández, A., 2017. Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Communication & Society,* 20(6), pp. 930-946.

Mayring, P., 2000. Qualitative Content Analysis. *FORUM: QUALITATIVE SOCIAL RESEARCH,* 1(2).

Mayring, P., 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution.* [Online] Available at: https://www.ssoar.info/ssoar/handle/document/39517 [Viewed 6 April 2019].

Merriam, S. B., 2009. *Qualitative Research: A Guide to Design and Implementation.* San Francisco: Wiley.

Milner, R. M., 2013. FCJ-156 Hacking the Social: Internet Memes, Identity Antagonism, and the Logic of Lulz. *The Fibreculture Journal,* 22, pp. 62-92.

Miškolci, J., Kováčová, L. & Rigová, E., 2018. Countering Hate Speech on Facebook: The Case of the Roma Minority in Slovakia. *Social Science Computer Review,* pp. 1-19.

Mouffe, C., 1999. Deliberative democracy or agonistic pluralism?. *Social Research,* 66(3), pp. 746-58.

Mouffe, C., 2005. *On the Political.* London: Routledge.

Neo Magazin Royal, 2018. *Hass im Internet | NEO MAGAZIN ROYALE mit Jan Böhmermann - ZDFneo.* [Online] Available at: https://www.youtube.com/watch?v=fAYjSLtz6wQ [Viewed 6 April 2019].

Neumayer, C. & Valtysson, B., 2013. Tweet against Nazis? Twitter, power, and networked publics in anti-fascist protests. *Journal of media and communication research,* 55, pp. 3-20.

Patton, M. Q., 2002. *Qualitative Research and Evaluation Methods.* Thousand Oaks: Sage.

Phillips, W., 2018. *The Oxygen of Amplification: Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online,* New York: Data & Society Research Institute.

Reconquista Internet, 2018. *RECONQUISTA INTERNET Mit Herz und Hand und mit Verstand.* [Online] Available at: https://www.facebook.com/pg/ReconquistaNetz/about/?ref=page_internal [Viewed 6 April 2019].

Reynolds, L. & Tuck, H., 2016. *THE COUNTER-NARRATIVE MONITORING & EVALUATION HANDBOOK,* London: Institute for Strategic Dialogue.

Richards, R. D. & Calvert, C., 2000. Counterspeech 2000: A new look at the old remedy for bad speech. *Brigham Young University Law Review,* 2, pp. 553-586.

Rieger, D., Schmitt, J. B. & Frischlich, L., 2018. Hate and counter-voices in the Internet: Introduction to the special issue. *Studies in Communication and Media,* 7(4), pp. 459-472.

Sängerlaub, A., Meier, M. & Rühl, W.-D., 2018. *Fakten statt Fakes. Verursacher, Verbreitungswege und Wirkungen von Fake News im Bundestagswahlkampf 2017,* Berlin: Stiftung Neue Verantwortung.

Schieb, C. & Preuss, M., 2016. Governing Hate Speech by Means of Counter Speech on Facebook.   [Online]   Available   at: https://www.researchgate.net/publication/303497937_Governing_hate_speech_by_mea ns_of_counterspeech_on_Facebook [Viewed 17 May 2019].

Schieb, C. & Preuss, M., 2018. Considering the Elaboration Likelihood Model for simulating hate and counter speech on Facebook. *Studies in Communication and Media,* 7(4), pp. 580-606.

Shifman, L., 2013. Memes in a Digital World: Reconciling with a Conceptual Troublemaker. *Journal of Computer-Mediated Communication,* 18(3), p. 362–77.

Sikorski, C. v. & Hänelt, M., 2016. Scandal 2.0 How Valenced Reader Comments Affect Recipients' Perception of Scandalized Individuals and the Journalistic Quality of Online News. *Journalism & Mass Communication Quarterly,* 93(3), pp. 551-571.

Silverman, T., Stewart, C. J., Amanullah, Z. & Birdwell, J., 2016. *THE IMPACT OF COUNTER-NARRATIVES: Insights from a year-long cross-platform pilot study of counter-narrative curation, targeting, evaluation and impact,* London: Institute for Strategic Dialogue.

Sponholz, L., 2016. Islamophobic Hate Speech: What is the Point of Counter-Speech? The Case of Oriana Fallaci and The Rage and the Pride. *Journal of Muslim Minority Affairs,* 36(4), pp. 502-522.

Teitelbaum, B. R., 2017. *Lions of the north: sounds of the new Nordic radical nationalism.* New York : Oxford University Press.

Thurston, N., 2019. Back to Front Truths. Hate Library. In: M. Fielitz & N. Thurston, Eds. *Post-Digital Cultures of the Far Right: Online Actions and Offline Consequences in Europe and the US.* Bielefeld: transcript, pp. 193-204.

Townsend, L. & Wallace, C., 2016. *Social Media Research: A Guide to Ethics.* [Online] Available at: https://www.gla.ac.uk/media/media_487729_en.pdf [Viewed 17 May 2019].

Uldam, J. & Kaun, A., 2019. Theorizing Civic Engagement and Social Media. In: M. Mortensen, C. Neumayer & T. Poell, Eds. *Social Media Materialities and Protest: Critical Reflections.* London & New York: Routledge, pp. 101-115.

Wendling, M., 2019. *BBC News.* [Online] Available at: https://www.bbc.com/news/blogs-trending-47596446 [Viewed 27 May 2019].

Zhang, Y. & Wildemuth, B. M., 2005. *Qualitative Analysis of Content.* [Online] Available at: https://www.ischool.utexas.edu/~yanz/Content_analysis.pdf [Viewed 6 April 2019].

Empirical sources

D-Generation, 2018. *HoGeSatzbau.* [Online] Available at: https://www.hogesatzbau.de/wp-content/uploads/2018/01/HANDBUCH-FÜR-MEDIENGUERILLAS.pdf [Viewed 25 May 2019].

Spiegel Online, 2019. *Studie: So denken die Deutschen über Migration.* [Online] Available at: https://www.facebook.com/spiegelonline/posts/eine-mehrheit-der-deutschen-sieht-einwanderung-als-chance/10157606649839869/ [Viewed 6 April 2019].

B.Z., 2019. *Die Tat zeigt, dass Rache und Hass mehr Rache und Hass hervorrufen.* [Online] Available at: https://www.facebook.com/B.Z.Berlin/posts/10157941251572437/ [Viewed 6 April 2019].

Reconquista Internet, 2019. *Reddit.* [Online] Available at: https://www.reddit.com/r/ReconquistaInternet/wiki/index#wiki_reconquistainternet_wiki [Viewed 6 April 2019].

Tag24 Dresden, 2019. *"MISSION LIFELINE"-KAPITÄN WIRBT FÜR MEHR MENSCHLICHKEIT.* [Online] Available at: https://www.facebook.com/tag24.dresden/posts/2554224704607621/ [Viewed 6 April 2019].

Appendix

Coding scheme: strategic communication documents

Main categories	Main codes	Sub-codes
Strategies and tactics		
Advocated values		
Target groups	Opponents Audience	

Coding scheme: user comments

Main categories	Main codes	Sub-codes
Type of comment	Digital Hate Culture	Hate Speech Misinformation Eristic Dialectic Hashtag
	Counter Speech	Expression of opposition Counter-argumentation Call for moderation Hashtag
	Page host	
	Meme/GIF	
Discourse	Islamophobia/anti-refugee Anti-establishment/anti-left Nationalism	