SCM
Studies in Communication and Media

FULL PAPER

Combatting hate and trolling with love and reason?
A qualitative analysis of the discursive antagonisms between organized hate speech and counterspeech online

Hass und Trolling mit Liebe und Vernunft bekämpfen?
Eine qualitative Analyse der diskursiven Antagonismen von organisierter Hate Speech und Counter Speech online

Nadine Keller & Tina Askanius

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Nadine Keller (M.A.) School of Arts and Communication, Nordenskiöldsgatan 1, 211 19 Malmö, Sweden. Contact: keller.ndn(at)gmail.com. ORCID: https://orcid.org/0000-0003-1879-370X

Tina Askanius (Ass. Prof. Dr.) School of Arts and Communication, Nordenskiöldsgatan 1, 211 19 Malmö, Sweden. Contact: tina.askanius(at)mau.se. ORCID: https://orcid.org/0000-0003-4953-2852

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

FULL PAPER

Combatting hate and trolling with love and reason?
A qualitative analysis of the discursive antagonisms between organized hate speech and counterspeech online

Hass und Trolling mit Liebe und Vernunft bekämpfen?
Eine qualitative Analyse der diskursiven Antagonismen von organisierter Hate Speech und Counter Speech online

Nadine Keller & Tina Askanius

Abstract: An increasingly organized culture of hate is flourishing in today's online spaces, posing a serious challenge for democratic societies. Our study seeks to unravel the workings of online hate on popular social media and assess the practices, potentialities, and limitations of organized counterspeech to stymie the spread of hate online. This article is based on a case study of an organized "troll army" of online hate speech in Germany, Reconquista Germanica, and the counterspeech initiative Reconquista Internet. Conducting a qualitative content analysis, we first unpack the strategies and stated intentions behind organized hate speech and counterspeech groups as articulated in their internal strategic documents. We then explore how and to what extent such strategies take shape in online media practices, focusing on the interplay between users spreading hate and users counterspeaking in the comment sections of German news articles on Facebook. The analysis draws on a multi-dimensional framework for studying social media engagement (Uldam & Kaun, 2019) with a focus on practices and discourses and turns to Mouffe's (2005) concepts of political antagonism and agonism to operationalize and deepen the discursive dimension. The study shows that the interactions between the two opposing camps are highly moralized, reflecting a post-political antagonistic battle between "good" and "evil" and showing limited signs of the potentials of counterspeech to foster productive agonism. The empirical data indicates that despite the promising intentions of rule-guided counterspeech, the counter efforts identified and scrutinized in this study predominantly fail to adhere to civic and moral standards and thus only spur on the destructive dynamics of digital hate culture.

Keywords: Online hate speech, counterspeech, Facebook, antagonism, trolling, far-right online ecosystem.

Zusammenfassung: In digitalen Räumen gedeiht aktuell eine zunehmend organisierte Kultur des Hasses, die eine ernsthafte Herausforderung für demokratische Gesellschaften darstellt. Unsere Studie erforscht die Praktiken von Online-Hass auf populären Social Media

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Plattformen und macht deutlich, welche Möglichkeiten und Grenzen organisierte Counter Speech (Gegenrede) hat, um der Verbreitung von Hass im Internet entgegenzuwirken. Der Artikel basiert auf einer Fallstudie über eine organisierte „Troll-Armee" des Online-Hasses in Deutschland, Reconquista Germanica, und die Counter Speech-Initiative Reconquista Internet. Anhand der internen strategischen Kommunikationsdokumente beider Gruppen führen wir eine qualitative Inhaltsanalyse durch um zunächst die Strategien und erklärten Absichten organisierter Hate- und Counter Speech-Gruppen zu ermitteln. Anschließend untersuchen wir inwiefern solche Strategien praktisch umgesetzt werden indem wir das Zusammenspiel zwischen NutzerInnen untersuchen, die in Kommentarspalten deutscher Nachrichtenartikel auf Facebook Hass verbreiten und denen, die mit Counter Speech entgegenhalten. Die Analyse orientiert sich an einem mehrdimensionalen Rahmen zur Untersuchung von Social Media Engagement (Uldam & Kaun, 2019) mit einem Schwerpunkt auf den Dimensionen Praktiken und Diskurse. Um letztere Dimension zu operationalisieren und zu vertiefen, stützt sich unsere Analyse auf Mouffes (2005) Konzepte des politischen Antagonismus und Agonismus. Die Studie zeigt, dass die Interaktionen zwischen den beiden gegnerischen Lagern stark moralisiert sind, was einen postpolitischen antagonistischen Kampf zwischen „Gut" und „Böse" zur Vorschau bringt, der nur in begrenztem Maße Anzeichen für das Potenzial von Counter Speech zeigt, produktiven Agonismus zu fördern. Die empirischen Daten deuten darauf hin, dass sich – trotz der vielversprechenden Praktiken von regelgesteuerter Counter Speech – die identifizierten und untersuchten Counter Speech Bemühungen überwiegend nicht an zivilgesellschaftliche und moralische Maßstäbe halten und auf diese Weise die destruktiven Dynamiken der digitalen Hasskultur nur weiter ankurbeln.

Schlagwörter: Online Hate Speech, Counter Speech, Facebook, Antagonismus, Trolling, Rechtsextremes Online-Ökosystem.

1. Introduction

Social media platforms have long been praised for their deliberative potential to provide people with spaces and possibilities to engage in rational and constructive political discussions (Friess & Eilders, 2015; Ruiz et al., 2011). Contrary to this anticipation, however, the comment sections on Facebook and other social media platforms have largely evolved into a *battleground* where internet users air their frustrations and spread hate for cultural "others" and those holding opposing views. Uncivil discourse and hate speech, understood as one of the most extreme forms of incivility (Lück & Nardi, 2019), saturate online discussions today, not least in Germany.[1] Along with the upsurge in populist far-right ideology across liberal democracies in recent years, and with it an increase in overt racist and Islamophobic attitudes, social media have, in fact, become a primary tool for both unorganized and organized dissemination of hatred to large audiences (Blaya, 2018). When online hate speech comes in organized, concerted forms, it is oftentimes orchestrated by so-called troll armies, some of which have disruptive and ideological agendas. Regardless of whether ideology is at the core of these online practices, hate speech is performative and not only has far-reaching conse-

1 A survey in Germany from 2018 shows that 78 % of Internet users have been confronted with hate speech online. Among the 14 to 24-year-old users, the share is 96 % (Landesanstalt für Medien NRW, 2018).

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

quences for the emotional and physical well-being of targeted groups but also poses a threat to pluralist democracy at large, promoting discrimination and intolerance and accelerating the polarization of societies (Rieger et al., 2018; George, 2015).[2] The definition of *hate speech* is highly contested and often used as a stand-in for several forms of group-focused enmity, such as antisemitism, racism, or hostility towards Sinti and Roma (Zick et al., 2016, p. 33-41). Indeed, as pointed out by Thurston (2019), hate speech is a deceptively complicated term, and various derivative and competing concepts are used to filter the kinds of expressions and intentions we might gather under such an umbrella term. For example, researchers in the Dangerous Speech Project propose that we distinguish between hate speech, which is hateful towards a group, and dangerous speech, which endagers that group because it inspires violence against them (Benesch et al., 2018). In this study, we follow Lepoutre's (2017) relatively broad definition of hate speech as "communications that emphatically deny the basic status of other members of society as free and equal citizens," thereby taking hate speech to "deny the basic standing of individuals who belong to vulnerable social groups, in virtue of their belonging to these groups" (p. 853). Importantly, hate speech does not necessarily come in verbal or written form but can also be conveyed visually, for instance, through GIFs or memes, as evidenced in the study at hand.

Although recognized as a pressing issue among legal institutions and policymakers (Awan, 2016), the current lack of efficient regulations against the workings of online hate speech makes counterspeech efforts and initiatives by individual internet users all the more important (Leonhard, Rueß, Obermaier, & Reinemann, 2018). As a response, individuals are increasingly joining forces online to counter the deluge of hate speech and to advocate a constructive discussion climate through various campaigns and online initiatives. This article examines the tug-of-war between these two camps of internet users – those intentionally spreading hate and those ostensibly practicing counterspeech in concerted online efforts – and seeks to understand the discursive antagonisms of their interplay as it unfolds on social media in order to assess the potential productiveness of organized counterspeech in stymieing the spread of hate speech. To do this, we draw on a case study of a recent confrontation in the German context between two such groups: Reconquista Germanica, an organized far-right troll army that covertly plots and executes cyberhate attacks against political enemies, and Reconquista Internet, a loosely organized civil society movement that aims to "bring reality back to the Internet" and combat the proliferating culture of digital hate with "love and reason" (Reconquista Internet, 2018). So far, there is little knowledge of the actual impact of such organized counter efforts; reliable empirical evidence of its effectiveness is scarce, just as existing case studies on the issue are mainly descriptive (Blaya, 2018). While scholars call for further research in the field, it has proven difficult to *measure* the effectiveness of user interventions to prevent and reduce online hate (Silverman, Stewart, Amanullah, & Bird-

2 Recent studies indicate that rising anti-Muslim hate speech on social media correlates with a surge of hate crimes against Muslims in both Germany and the UK (Müller & Schwarz, 2017; Awan & Zempi, 2017).

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

well, 2016). We take a qualitative approach to detecting the patterns characterizing the practices of digitally organized hate speech and counterspeech groups, which allows us to understand their various manifestations, nuances, and related (in)effectiveness. We direct detailed analytical attention both to the intentions of two concrete groups of hate speech and counterspeech and to online discussions where these opposing forces collide. This enables us to assess which communicative practices are conducive for the strategic action against incivility and hatred online and which efforts are counterproductive and even play into the hands of those stirring up hatred.

The analysis proceeds in two parts. We first conduct a qualitative content analysis of a set of internal strategic communication documents from both groups, exploring their stated intentions, strategies, and tactics. Second, we turn to how such intentions find expression on social media by analyzing user discussions in Facebook comment sections appertaining to news articles on topics related to immigration, refugees, or Islam. In Germany, the public debate on these topics intensified after Chancellor Merkel's optimistic "Wir schaffen das!" [We can do it!] response to the refugee question in 2015, which has since continued to polarize political debate in the country. Therefore, and because anti-Muslim sentiment is still a nexus of for the diverse groups constituting the culture of digital hate internationally (e.g., Davey & Ebner, 2017; Jaki & De Smedt, 2018), we chose to analyze discussions on these themes.

Theoretically, this paper draws on Mouffe's (2005) work on the antagonistic nature of "the political" and today's post-political Zeitgeist, in which she takes issue with the idea of consensus-oriented deliberation and argues that democracy *needs* passionate, even heated, debates to thrive. Her understanding of democratic participation as essentially conflict-ridden and emotional allows us to approach the partisan confrontations on social media we see in our data not (only) as expressions of divisive antagonisms that are harmful to society but potentially as the productive agonistic pluralism that is vital for a healthy and functioning democracy. Further, our research is inspired by the multi-dimensional framework for studying civic engagement on social media proposed by Uldam and Kaun (2019), which offers a holistic model paying attention to the (i) practices, (ii) discourses, (iii) power relations, and (iv) technological affordances shaping civic engagement online. To limit and operationalize this case study, we focus on the analytical dimensions of *practices* and *discourses* as they allow us to acknowledge the constitutive and performative power of user interactions. With these objectives in mind, this article sets out to answer the following research questions: (1) What characterizes the communicative strategies and commenting practices of organized digital hate culture and organized counterspeech? (2) How are antagonisms discursively constructed when hate speech and counterspeech efforts clash on Facebook? (3) How productive are organized counterspeech efforts in stymieing the spread of hate speech online and fostering productive agonism?

With the growing body of research on the proliferation of organized hatred throughout the Internet and the relative lack of research on effective ways to counter it, this study joins a timely and controversial conversation. Considering the overwhelming bias in current research on "hate politics" towards US and UK

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

contexts (Benkler, Faris, & Roberts, 2018), this study makes a specific contribution to this field by examining how these global conflicts play out in the current political climate in Germany post the European border crisis of 2015–16. The produced knowledge can potentially enhance and improve organized counter efforts, thereby effectively undermining the destructive dynamics that currently thrive in online spaces.

2. Online practices of and responses to digital hate culture

The following theoretical reflections on key terms and concepts shed light on the complex nature of far-right online ecosystems and help us draw a clear line between organized and politically motivated online hate and the idea of "trolling" as "harmless fun" that Reconquista Germanica wrongfully brands itself with. We further look at potential pros and cons of counterspeech in its different forms of expression, which substantiates our endeavor to explore the promising potentials of Reconquista Internet and also allows us to understand the drawbacks of counter efforts that do not adhere to civic and moral standards.

2.1 A hodgepodge of haters and their shared practices of manipulation

It is a common 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 specific or clear-cut actors on the far right, we use the term *digital hate culture* (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. Proponents of digital hate culture who openly admit to a political or activist agenda cover a wide spectrum of far-right positions and exhibit different levels of extremism and antidemocratic sensibility. Mapping the complicated mesh of actors that make up the burgeoning and highly adaptable new far-right across Europe and the United States, Davey and Ebner (2017, p. 7) find that, despite their different ideological differences and varying degrees of extremism, many of these groups converge and actively collaborate internationally on achieving common goals – most significantly preventing immigration, removing hate speech laws, and bringing far-right politicians into power. Their collaboration is undergirded and facilitated by the appropriation of Internet culture and shared online communication strategies across an alternative ecosystem of online media (Marwick & Lewis, 2017; Davey & Ebner, 2017). Ebner (2019) describes this online influence ecosystem as circular: A wide range of so-called Alt-Tech platforms (e.g., chatrooms and forums) with lax community policies and strong advocacy of free speech, like 4chan and Gab, often referred to as fringe communities, allow far-right groups to mobilize action and recruit new members. The coordination of action and further radicalization then moves to encrypted messaging apps, like Discord. Finally, these groups move from the fringe communities and make concerted efforts to distort

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

public discourse in popular social media platforms like Facebook and the comment sections of news articles circulating there. Hate speech is the pivotal form of expression of digital hate culture and should be intimately understood in relation to the rise of these alternative online ecosystems. At the core of existing definitions of hate speech is the emphasis on an identity-based enmity towards certain groups. Typically, hate speech exploits a fear of the unknown based on identity categories like race, ethnicity, religion, and gender (Leonhard, Rueß, Obermaier, & Reinemann, 2018). However, the scapegoating can also be directed against the establishment, its institutions, and its proponents (*Systemlinge*) – including the mainstream media (*Lügenpresse*), left intellectuals, and general "social justice warriors" (*Gutmenschen*) (Darmstadt, Prinz, & Saal, 2019). 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, Schmitt, & Frischlich, 2018).

More than overt hate speech, the mainstreaming practices of digital hate culture rely on exploiting visual and aesthetic means that help to promote ideology in subtle but persuasive ways and that can often circumvent censorship (Bogerts & Fielitz, 2019). This is most apparent in the wide appropriation of memes, which are images "that quickly convey humor or political thought, meant to be shared on social media" (Marwick & Lewis, 2017, p. 36). Co-opting these popular media forms from both mainstream digital cultures and online fringe communities offers activists a gateway for expressing radical ideas in an easily digestible, seemingly harmless, and "funny" way. Humor has become an essential part of the mainstreaming practices characteristic of digital hate culture (Askanius, 2021; Schwarzenegger & Wagner, 2018). Meanwhile, the slippery term *trolling* is used to describe the wide range of asocial online behavior that appropriates humor to sow discord and provoke emotional responses (Marwick & Lewis, 2017). By claiming 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). We bring all of these concepts into the mix to promote an understanding of Reconquista Germanica as essentially a troll army embracing and practicing organized hate speech; concurrently, we address the communication strategies applied across the digitally-driven far-right media ecosystem that use memes, humor, and irony as a way of weaponizing contemporary internet culture, with all of its humorous ambiguities, to promote far-right political agendas.

2.2 Regulatory and civil society responses – the role of counterspeech

In Germany, policymakers and legal actors are aware of the danger digital hate culture poses to society and have proposed regulatory measures, such as the Network Enforcement Act (short NetzDG). The law forces social media companies to implement procedures that allow users to report illegal or extremist content and obliges the platforms to canvass user complaints immediately and to delete content "obviously punishable by law" within 24 hours (Darmstadt, Prinz, & Saal, 2019). Although such regulatory measures constitute a novel and aggressive ap-

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

proach to crack down on digital hate culture,[3] procedures to repress content are also criticized since they raise difficult questions about the limits of free speech and the definition of hate (Rieger, Schmitt, & Frischlich, 2018). The impending high fines in case of infringement are likely to lead to an *overblocking* of content that is not unlawful; consequently, public perception of 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 are not punishable by law: Exponents of digital hate culture often operate *strategically*, being well aware of legal boundaries, and apply measures, such as humor, to circumvent them. Hence, many NGOs and civil society initiatives call for alternative means to fight digital hate culture, such as organized counterspeech.

Scholarly definitions of counterspeech are scarce. In our reading of the term, we draw mainly on work presented in a special issue on the topic in this journal (see Rieger, Schmitt & Frischlich, 2018). Accordingly, we understand counterspeech to encompass all communicative actions aimed at refuting hate speech and supporting targets and fellow counter speakers through thoughtful, cogent reasons and true fact-bound arguments (Schieb & Preuss, 2018). Such forms of counterspeech may contribute to "civic education in a broader sense" by "disseminat[ing] messages of tolerance and civility" (Rieger, Schmitt, & Frischlich, 2018, p. 464). Schieb and Preuss (2018) found that counterspeech achieves maximum impact if it is organized, conducted in groups, quick in reaction, and not too extreme in its positions. Although regarded as one of the most important responses to hate speech, counterspeech also has its downsides. On the one hand, the act of confronting online aggression can result in negative or even dangerous experiences for the counter speaker (Blaya, 2018). On the other hand, counterspeech can provoke unwanted side effects when counter speakers do not adhere to civic and moral standards themselves (Rieger, Schmitt, & Frischlich, 2018); consequently, it risks providing hateful content with relevance, "discussability," and thus better discourse quality (Sponholz, 2016). Essentially, organized practices of counterspeech, like the ones explored in this article, are based on the assumption that we may successfully counter hate speech with more (democratic) speech. Instead of coercively excluding or banning hate speech, even when it poses a danger to certain social groups, we should tolerate such acts as part of the inherent agonism of pluralist society and address them as part of the tedious but necessary "democratic discursive processes" (Lepoutre, 2017, p. 855).

2.3 Case study: Reconquista Germanica vs. Reconquista Internet

Calling itself the "biggest patriotic German-language Discord channel" (Davey & Ebner, 2017), Reconquista Germanica's (RG) primary goal is to reclaim cyberspace, influence online discourse, and exert pressure on politics (Kreißel, Ebner, Urban, & Guhl, 2018). The case of RG demonstrates that trolling is not just a harmless practice of apolitical subcultures who are after the "lulz" but also a practice appropriated by users with a clear political agenda: While claiming to be a mere "satirical project that has no connection to the real world" (Anders,

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access -

2017), RG is believed to have had a crucial influence on the German parliamentary elections in 2017 in pushing forward right-wing populist party AfD (Kreißel, Ebner, Urban, & Guhl, 2018; Guhl, Ebner, & Rau, 2020). The group counted about 5,000 members willing to invade the Internet with coordinated raids to polarize and distort public discourse on the day before the general election (Bogerts & Fielitz, 2019). The Discord server was set up by a German far-right activist operating under the pseudonym Nikolai Alexander, whose YouTube channel counts over 33,000 subscribers. RG assembles "patriotic forces" from a wide spectrum of cultural nationalists; loosely organized Neo-Nazis; far-right influencers; and members of the Identitarian Movement and the right-wing parties AfD, Junge Alternative, and NPD (Bogerts & Fielitz, 2019; Ayyadi, 2018). The platform is dominated by racist, xenophobic, and National Socialist content and symbols, showcasing a mixture of military language derived from the Nazi-German Wehrmacht and alt-right insider vocabulary (Kreißel, Ebner, Urban, & Guhl, 2018; Davey & Ebner, 2017).

At the end of April 2018, shortly after the disclosure of RG by a group of journalists, satirist and moderator Jan Böhmermann initiated Reconquista Internet (RI) during the popular German satire news show *Neo Magazin Royal*.[4] Initially meant as a satirical reaction, RI grew into one of the biggest civil society-led counterspeech initiatives in Germany. Böhmermann's call for a "love army" to fight back against the trolls and reconquer the Internet attracted a large number of supporters, reaching 60,000 people today. RI is a digital, non-partisan, and independent civil rights movement that describes itself as a "voluntary and private association of people of all ages, genders, and backgrounds and with all sorts of other individual benefits and mistakes" (Reconquista Internet, 2018). Everyone who respects the German Basic Law and their ten-point codex can join the network. 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). Similar to 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 counterspeech. Due to the high popularity of Jan Böhmermann and his satire show in Germany, RG and RI are likely the most known and visible groups of their kind in the national context, making them a suitable subject for this study.

3. The antagonistic nature of the political and the potential of productive agonism

This study is informed by Chantal Mouffe's (2005) critique of today's post-political Zeitgeist and her theoretical perspectives on the political as a dimension of antagonism that can take many forms and can emerge in diverse social relations. Her work on the political in its antagonistic dimension helps us understand the interplay between (1) the anti-democratic sentiments and behaviors flourishing in

4 The show was discontinued in December 2019.

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

digital hate culture and (2) the democratic forms of civic engagement seeking to combat hate and bigotry at the heart of organized counterspeech. According to Mouffe (2005; 2016), all political struggle should be understood in terms of discursive articulations of political identities. Political identities are relational and collective and always entail a "we/they" discrimination. As antagonism is constitutive of human societies, it is the role of democratic *politics* to create order. According to Mouffe (2005) , 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. (p. 20)*

In order to prevent we/they relations from devolving into friend/enemy relations, Mouffe sees the main task of democracy as providing space for a plurality of adversarial ideological positions and ensuring them legitimate forms of expression. In one of her most recent essays, she argues,

> *When dealing with political identities ..., we are dealing with the creation of an "us" that can only exist by its demarcation from a "them." This does not mean of course that such a relation is by necessity an antagonistic one. But it means that there is always the possibility of this us/them relation becoming a friend/enemy relation. (Mouffe, 2016)*

We consider her idea of antagonism as an "ever-present possibility" inherent to all social interaction between opposing parties and the impetus to turn enemies into adversaries as the driving force behind organized counterspeech initiatives like the one addressed in this study.

The relative success of right-wing ideology across Germany and the rest of Europe in recent years can partly be understood as a consequence of the blurring of boundaries between left and right as meaningful political categories with which to orient our democratic engagement and frustrations. In this post-political era, politics is increasingly played out in a moral register in which the we/they discrimination – 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, Mouffe ascertains. She argues that "when the channels are not available through which conflicts could take an 'agonistic' form, those conflicts tend to emerge on the antagonistic mode" (Mouffe, 2016, p. 5). Instead of trying to negotiate a compromise or rational solution among the competing camps, Mouffe proposes a radicalization of modern democracy and advocates agonistic pluralism, in which conflicts are provided with a legitimate form of expression.

However, when investigating participatory online media's potential to allow antagonisms to be transformed into agonistic modes of expression that are beneficial for democracy, scholars have come to different conclusions. For example, Milner (2013) highlights the potential of the irony-laden communication prac-

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access -

tices underlying the "logic of lulz" to be adopted by counter publics in order to encourage a vibrant, agonistic discussion over exclusionary antagonism. Other scholars point to the limitations of Mouffe's argument when talking about digital hate culture and critically question "whether being a racist is a democratic right [and] whether freedom of speech includes opinions and views that challenge basic democratic values" (Cammaerts, 2009, p. 1). When discussing the extent to which democratic tolerance for difference can offer a plausible response to hate speech, Davids (2018) 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. Our study follows Mouffe's understanding of politics as a discursive battlefield where competing groups with opposing interests vie for hegemony. The online clash between exponents of digital hate culture and counter speakers showcases the alarming level of antagonism currently flourishing in German society and beyond. In the analysis of such confrontations, we draw on Mouffe's theory to understand the democratic potential inherent to conflict as it unfolds among citizens on social media. Her perspectives help us reflect on the limits of regarding counterspeech as universally productive means to address and potentially bridge the antagonism inherent to all human societies; simultaneously, they allow us to envision democratic politics in new, more agonistic ways – as we will see that civil society-led counterspeech as promoted by organized groups does indeed give hope for transforming antagonism into agonism.

4. Research design

The analytical framework guiding this research is inspired by Uldam and Kaun's (2019) four-dimensional model for studying civic engagement on social media. The model highlights the importance of moving beyond a media-centric perspective to understand user interaction on social media as a holistic ensemble, determined by the four dimensions (i) practices, (ii) discourses, (iii) power relations, and (iv) technological affordances. While we focus our analysis on the first two dimensions of intended and actualized practices (RQ 1) and embedded discourses (RQ 2), we position these in relation to questions of the underlying structures (such as algorithms, regulation, and ownership) in order to understand the broader context and techno-political infrastructures that impinge on the online spaces in which citizens engage in discussions about contentious political issues today. The insights gained eventually allow us to critically reflect on the potentials of organized counterspeech efforts (RQ 3) and indicate fruitful directions for future research.

4.1 Method, empirical material, and selection criteria

The analysis draws on a qualitative content analysis, which allows us to detect themes and establish patterns across media texts and representations that illustrate the range of meanings and forms of expression of our research phenomenon (Mayring, 2014; Krippendorff, 2004). To meet the demands of scientific rigor, the description and subjective interpretation of data follow the systematic classifica-

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

tion process of coding. The analysis proceeds in two steps. The first dataset, comprising strategic communication documents of both groups, is analyzed following Mayring's (2014) summary of content, which helps reduce the material to its core content by paraphrasing and generalizing the data. In the case of RG, the sample consists of the ten-page document *Handbook for Media Guerillas*, which comprises four parts of how-to guidelines for (1) shitposting, (2) open-source memetic warfare, (3) social networking raids, and (4) attacking filter bubbles (HoGeSatzbau, 2018). Originally circulated on the far-right website *D-Generation* in May 2017, the document was later leaked from RG's Discord server and made public by the counterspeech-initiative HoGeSatzbau (Hooligans Against Syntax) in January 2018 (Lauer, 2018). Having public access to strategic documents that are usually sealed off in a closed group offers a unique opportunity to gain insights into the organized operating and intended practices formative of digital hate culture. In the case of RI, the sample comprises the group's wiki page on Reddit, which contains all project-relevant information as well as the group's ten-point-codex, guides, and instructions (Reconquista Internet, 2019).

The second dataset, which consists of user comments from three Facebook comment sections, is analyzed following Mayring's structuring of content, in which aspects of the material are filtered out based on predetermined criteria. This dataset provides a window into *actualized* practices of haters and counter speakers on social media. With research showing that immigrants, specifically Muslim refugees, are the group most targeted by online hate in Germany (Jaki & De Smedt, 2018), we purposefully sampled user discussions that demonstrate thematic relation to the polarizing topics of Muslim immigration and Islam. To identify a sample that was likely to showcase an active involvement of organized counter speakers as well as an engagement of digital hate culture exponents, we followed the coordinated counteracts of RI. However, once the discursive battle between hate speech and counterspeech has moved from fringe communities to mainstream social media like Facebook, clear indicators of group affiliation are scarce.[5] To avoid this becoming a potential threat to the internal validity of the study, we had to engage in extensive empirical legwork in the first step of analysis to familiarize ourselves with the digital footprints left by organized hate speech and counterspeech groups. Instead of trying to allocate discussion participants to either movement (while we can assume RI's presence, there can be no certainty that RG was involved), we used the second dataset to bring the clear intentions identified through our case study and the diffused intermixed practices found on popular social media into dialogue and to scrutinize the various forms that hate speech and counterspeech can take.

Once verified for RI's Discord server, we gained access to different workshop areas, including WS Counterspeech. This workshop is for members to post links to discussions on different social media platforms where they have detected hateful

5 RI consciously avoids clear indicators (such as hashtags) that would allow content to be allocated to an organized counterspeech group because an accumulation of such provides patters easily recognizable for algorithms, which exposes the group to the risk of systematic blocking (Reconquista Internet, 2019).

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

content and call for fellow members to counteract. Following the links posted in the sub-chat for Facebook at the time of data collection (February to March 2019), we sampled user discussions belonging to three articles of German news outlets.

Table 1. Selected articles

Title	Content	Date posted on Facebook	News outlet	Number of comments
„Mission Lifeline"-Kapitän wirbt für mehr Menschlichkeit [Captain of "Mission Lifeline" promotes greater humanity]	A report on Claus-Peter Reisch, captain of the rescue ship *Lifeline*, who advocated sea rescue operations at an event in Dresden.	24 February 2019	Tag24 Dresden	139
Die Tat zeigt, dass Rache und Hass mehr Rache und Hass hervorrufen [The act shows that revenge and hate create more revenge and hate]	A commentary on the right-wing extremist terrorist attack in Christchurch, New Zealand, condemning the concepts of hate and revenge.	16 March 2019	B.Z.	105
Studie: So denken die Deutschen über Migration [Study: This is how Germans think about migration]	A report on a study that shows that a majority of Germans (53 %) sees immigration as an opportunity.	19 March 2019	Spiegel Online	465

Among a large number of links shared by RI members, these three articles were identified as suitable samples because they were the only ones that met both predefined selection criteria – the topical focus on Muslim immigration and/or Islam and a minimum user engagement of >100 comments at the time of data collection. While *Tag24 Dresden* and *B.Z.* are regional news outlets (with a Facebook fan base of 100,000 to 135,000), *Spiegel Online* (with close to 1,9 million Facebook fans) is one of the most widely read German-language news websites; therefore, it has a large number of comments and considerable content moderation practices in place (Boberg et al., 2018). To limit the impact that comment deletion by content moderators might have on the sample while simultaneously allowing for sufficient user engagement, all comments were retrieved within 24 hours after publication. By recording "All Comments" instead of Facebook's default display of "Most Relevant" comments, the data collection resulted in a final sample of 709 initial user comments and all their reply comments (which on average included four per comment) in the form of text and visuals (emojis, GIFs, and memes).

4.2 Coding procedure

All documents and comment threads were organized and coded employing NVivo software. Tailored to the two focal areas of the analytical framework (practices and discourse), we developed and derived themes in form of categories, both in-

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

ductively from the data itself and deductively from the theoretical concepts and ideas introduced above. Combining these approaches allowed us to stay attentive to the data's individual properties and, at the same time, account for the extensive research in the field during the past years. We identified relevant main categories and subcategories that were complimented throughout the coding process. The first dataset was coded based on the identification of three main categories: *strategies and tactics* (overall goals and concrete means to achieve those), *advocated values* (principles that guide the groups' actions), and *target group(s)* (intended audiences). The coding scheme for the second dataset was aimed at determining the *type of comment* (hate speech, counterspeech, follower) and *discourse* (the presence of Islamophobic/anti-refugee, anti-establishment/anti-left, or right-wing populist/nationalistic sentiment). For comments that were identified to contain either hate or counterspeech, we further coded means of reaction (the number of symbolic reactions [such as Facebook's "like," "love," etc.] and the (un)willingness of respondents in reply comments to seek mutual understanding). The coding was carried out by the two authors of this article, first, on a sample to establish intercoder reliability (resulting in a mean agreement of 91%[6]) and to validate coding consistency and, subsequently, on the entire corpus of text, including the whole dataset of strategic communication documents and all collected user comments with coding units varying between single words, sentences, paragraphs, or entire user comments.

5. Analysis: Discursive antagonisms in the disjuncture between intentions and practices

The analysis showed a strong analogy between the first dataset on the side of RG and the second dataset of hate speech identified in the Facebook discussions. On the other hand, the intentions of organized counterspeech, as advocated by RI, and the counterspeech efforts found in the comment sections display great disparities. Acknowledging this gap between thoughtful, fact-bound counterspeech and attempts of refuting hate speech by adopting norm-transgressing practices similar to those characteristic of digital hate culture is critical for making a nuanced assessment of the agonistic potential counterspeech holds.

5.1 Stated intentions: Aggressive "warfare" vs. a defense strategy of "respect, love, and reason"

The analysis of RG's and RI's internal documents reveals how organized groups of digital hate culture and counter initiatives operate highly *strategically* to achieve their respective goals. 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 focuses on *followers*

6 The mean agreement relates to the analysis of user comments only, and the intercoder reliability test was conducted based on a smaller sample size representing 10% of the total population of comments.

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

(Schieb & Preuss, 2018), that is, the undecided audience. For the former target group, RG's *Handbook for Media Guerillas* advises infiltrating 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" (HoGeSatzbau, 2018). Strategies to "fight the enemy" rely on humiliation, provocation, and discreditation. Tactics include tagging "lies" as #fakenews, creating fake accounts in the enemy's name, drawing on a prepared repertoire of insults, swinging the "Nazi club" to turn the enemy's accusations of racism and anti-Semitism around, and always having the last word, all the while feigning calmness and politeness to make the opponent livid with rage. RG understands its opponents as enemies, not adversaries, hence mobilizing its proponents to fight using all means necessary. Contrary to this approach of intended offense, RI promotes a communication strategy of defense, as epitomized in the first three 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 (Reconquista Internet, 2019)*

RI rejects top-down instructions and instead recommends counter speakers to search for common grounds on which both discussion participants can build their conversation. The group works towards *understanding* its opponents, following the tactics of listening actively, asking open questions, switching perspectives, preventing thematic changes, motivating own opinions, sticking to objective criticism, and staying fair and calm, while always interpreting the opponent's arguments in the most favorable way possible. Such practices work towards a reduction of exclusionary antagonism by accepting the opponent's standpoint as legitimate and by rejecting moral assessments of "right" and "wrong" and polarizing "us/them" distinctions.

Interestingly, the analysis indicates that the intended practices of both groups are not focused on provoking/resolving conflicts with the deadlocked positions of the counterparty but mainly aimed at affecting followers, namely, users who enter or observe discussions without positioning themselves in either one of the two opposing camps. To do so, RI relies on knowledge-sharing and objective, factual reasoning, whereas RG focuses on obfuscation, argumentative manipulation, and emotionalizing content: "Note: In online discussions, 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" (HoGeSatzbau, 2018). Moreover, the *Handbook for Media Guerillas* advises "shitposters" to be friendly and to extend a "helping hand" to those audiences who are willing to learn the "truth" (Ibid.). Hashtags, such as #wakeup, are used to target those still wavering between different opinions. Meanwhile, RI builds on an extensive knowledge basis about these often-veiled strategies and tactics of troll armies and works to uncover them so that other discussion participants, as well as

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

silent audiences, realize the tricks that are used to tamper with the art of argumentation.

Both groups stress the importance of using humor and of operating in teams. Postulating that "a laughing enemy is already halfway on our side," the *Handbook for Media Guerillas* calls on its adherents to be funny and creative in their communication. RG appropriates the "logic of lulz" to veil its actions as being "merely for fun" (Milner, 2013). Branding the handbook as a "manual for mockery" trivializes the advocated systematic manipulation of public opinion (HoGeSatzbau, 2018). In contrast, RI poses a clear set of rules for what can and cannot be joked about: for example, memes should never contain swear words, insults, defamations, or violent references and must never target single individuals. For the counter speakers, teamwork means mutual support through endorsing each other's comments or "upvoting" content through, for example, Facebook's "like" and "love" reactions; in contrast, for RG, it serves as a means to exploit algorithmic processes and fight so-called information wars: Members are mobilized on encrypted messaging platforms to simultaneously infiltrate popular social media with multiple inconspicuous fake accounts and coordinated hashtags that make their content trending.

5.2 Actualized practices:The antagonistic battle between "good" and "evil"

When moving from the stated intentions of the two groups to the more messy realities of how these efforts are actually carried out and practiced on Facebook, the picture becomes more complicated. The results of our analysis show that around 40 percent of user comments contain forms of hate speech or counterspeech. The remaining 60 percent mainly feature forms of self-positioning and opinion-stating that do not fall under our definitions of hate speech and counterspeech but still show signs of affiliation to one of the opposing sides (e.g., when users accuse each other of stirring up hatred without drawing on or defending against identity-based enmity). There are barely any instances indicating that a user is uncertain or looks to challenge or change her or his opinion. Considering the objectives of both sides to primarily address undecided followers and the fact that these susceptible users appear to be least engaged in the discussions, the commenting efforts come down to counter-arguing the opponent's statements in a way that the *silent* audiences may be convinced of the truthfulness of one of the opposing standpoints. This dynamic intensifies the need to vehemently defend the opposing standpoints and *defeat* the opponent by all means, leaving little room for the promising agonistic approach RI stipulates.

5.2.1 The tug-of-war for the "right" representation of immigration, Islam, and the "deceitful system"

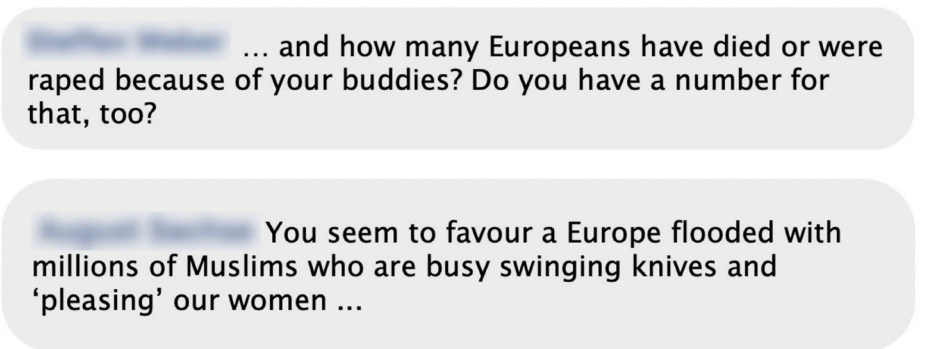
The discussions between the two opposing camps essentially revolve around different articulations of the "truth" that defends the "right" representation of immigration, Islam, and the role that "higher structures" play in this. Whereas one side promotes discourses of hostility and exclusion, the other side calls on human

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access -

values and advocates equality, tolerance, and openness. However, only few counter speakers follow the strategies and tactics promoted by RI and adhere to standards of respectful, true, fact-bound, and thoughtful reasoning. The majority of counter efforts act in obtrusive and authoritarian ways (i.e., telling users how to behave), and some make use of similar aggressive and derogatory language to defend verbal and visual attacks on refugees and Muslims as those disseminating identity-based hatred. This greatly impairs the potential of such forms of what we might call *spontaneous counterspeech* to encourage a productive agonistic discussion climate.

We identified a wide range of negative representations of Muslims, refugees, and Islam (Figure 1) that illustrates how digital hate culture works to generalize and denigrate entire groups of people, using "threats, discrimination, intimidation, marginalizing, otherings and dehumanising narratives" (Blaya, 2018, p. 2).

Figure 1. Recurring allegations against Muslims and refugees

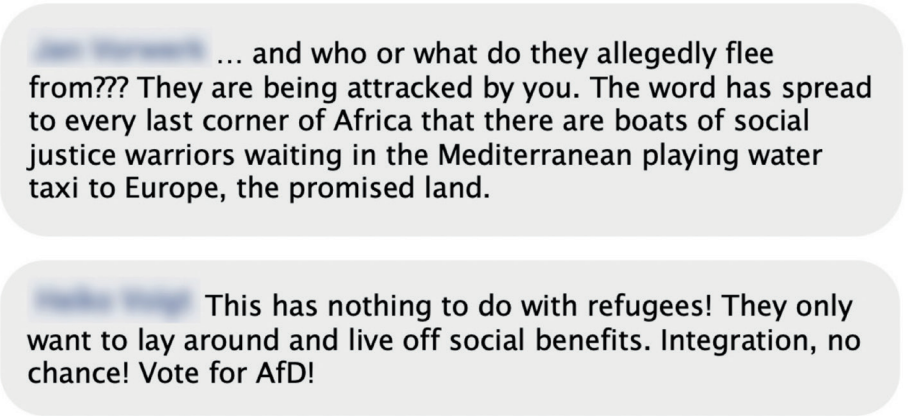


Note. Translated and visually reconstructed user comments.

The Islamophobic discourse present in the discussions resembles the themes identified in anti-Muslim hate speech: the rejection of refugees; concerns about a cultural displacement due to an "Islamic invasion"; and the threat of terrorism, (sexual) violence, and misogyny (Davey & Ebner, 2017). The discourse that is perpetuated through such representations antagonizes Muslim immigrants as the cultural other and creates a binary opposition between Islam and Christianity that suggests the incompatibility of Muslims with Western culture. The spread of disinformation, a key manipulation practice of digital hate culture (Marwick & Lewis, 2017), was a recurrent tactic in the analyzed comment sections. The discussion about sea rescue missions for refugees is pervaded by comments suggesting that the majority of immigrants who came to Europe since 2015 do not classify as actual refugees. A popular explanation for people's "true" motivation to leave their homes is the appeal of the European social welfare systems. This allegation is underpinned by the conspiracy theory that Europe is, in fact, luring people to come here since governments benefit from the booming asylum industry (Fig. 2).

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 2. Conspiracy theory about the "true" motivation of people to come to Europe



Note. Translated and visually reconstructed user comments.

Conspiracy theories are interwoven to make a sound argument for the reader and cultivate a "common sense" (Ganesh, 2018). The overarching themes in such conspiracy theories express anxieties about losing control within a political, religious, or social order (Marwick & Lewis, 2017). The dominant anti-establishment discourse, with allegations made against the "deceitful system," often spins around the "White genocide" and "Red Pill"[7] tropes that are crucial points for the fractured groups of digital hate culture. Any efforts to counter such narratives, whether with abusive language or fact-bound reasoning, are turned around to fortify the claim of speaking a truth that is being suppressed by "the leftist system." To unknowing audiences, such supposed insights can appear to reveal a hidden truth that governments are withholding, making it a powerful tactic to convert others to totalitarian and extreme worldviews. Accepting any such presuppositions makes it easier to incite hatred and more acceptable to not show empathy for the hardships of the dehumanized other. Memes are used strategically to present such allegations in simplified and affecting ways, increasing their chances to appeal to "multiple audiences far beyond those who unambiguously identify with neo-Nazi and other far-right symbolism" (Bogerts & Fielitz, 2019, p. 150); see Figure 3 as an example.

7 The "White genocide" conspiracy theory contends that Western civilization faces an existential threat from mass immigration and racial integration that is believed to be a deliberate action aiming to dismantle white collective power. To take the "Red Pill" means to become enlightened to the deceit committed by the "leftist project" (referring to feminists, Marxists, socialists, and liberals) who conspire to destroy Western civilization and culture (Ganesh, 2018).

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

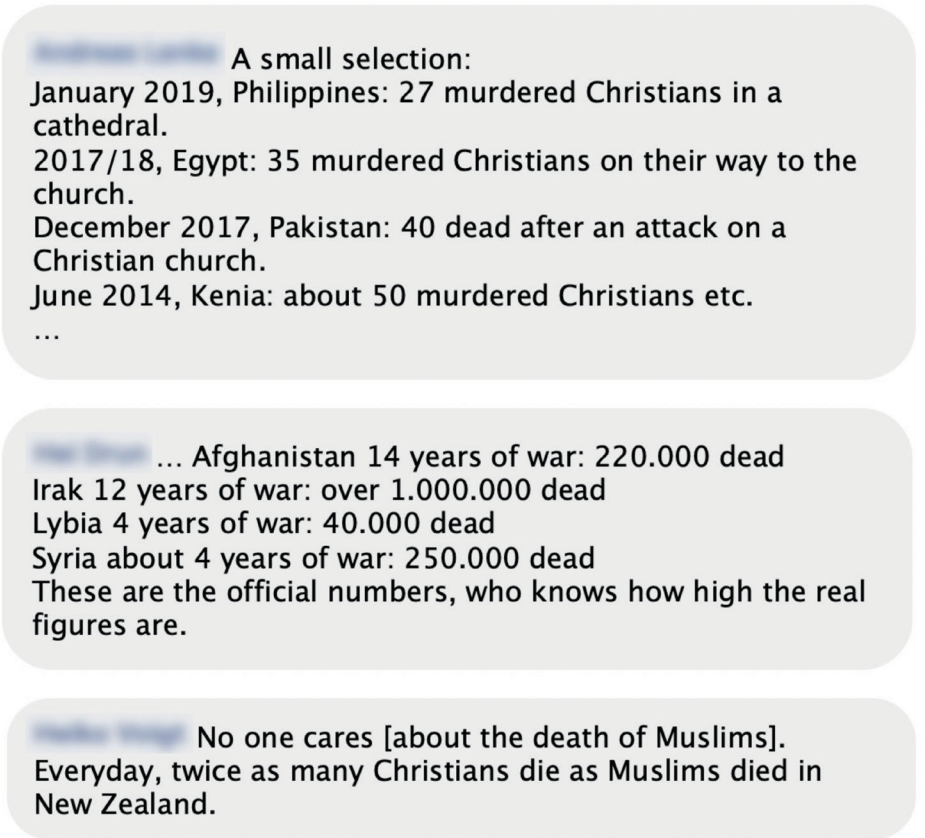
Figure 3. Anti-establishment and anti-refugee meme

Note. Translation: "As long as you bring them into our country instead of them, you hypocrites can kiss my ass!"

All analyzed comment sections display a strong moralization of Islam and the West/Christianity, juxtaposing them as "evil" and "good," respectively. People are repeatedly being generalized based on their supposed religious identity and blamed for past events in which the respective religion has endorsed violence. Interweaving several de facto and unconnected narratives creates what Darmstadt et al. (2019) call "toxic narratives" that are used to relativize or enforce the significance of single events. By recurrently postulating correlations and causalities, such narratives "stir up emotions and can help to motivate and mobilize" (Ibid., p. 160), making them valuable tools for sowing hatred and fear. In the example below (Figure 4), such toxic narratives evolved into a form of competition, in which users from both camps try to outdo each other with numbers of victims that can be ascribed to the wrong-doing of either Islam or Christianity.

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 4. Toxic narrative weighing the value of human lives according to religion

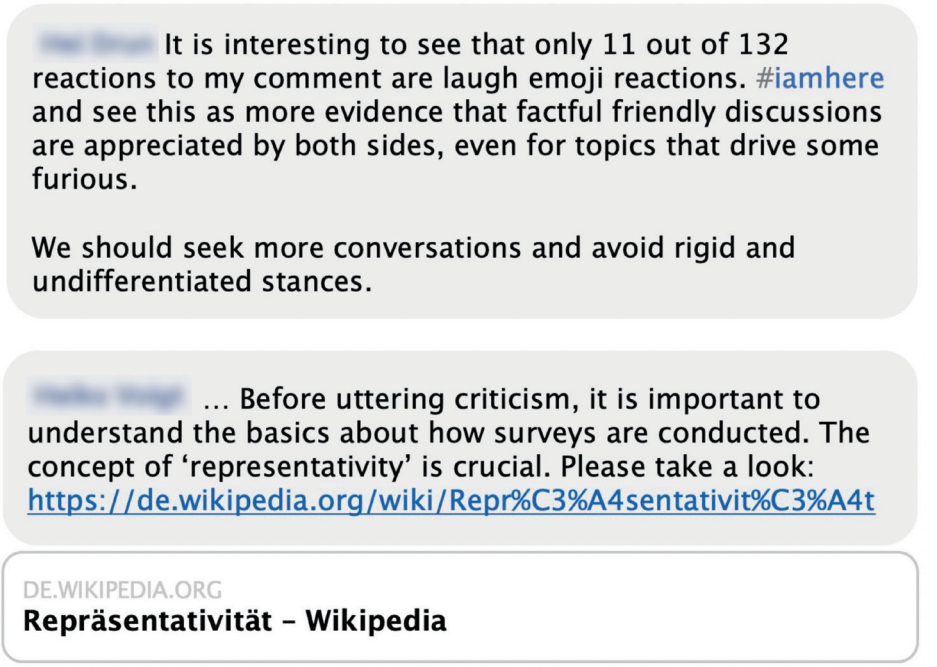


Note. Translated and visually reconstructed user interaction from the comment section on the Christchurch attacks.

This example reflects how, in the battlefield of digital hate culture and counterspeech (as it predominantly manifests in the analyzed discussions), the political is played out in a *moral* register, which establishes we/they distinctions in terms of "right" and "wrong." The fact that neither side sees the standpoint of the opponent as legitimate or encourages a communication-oriented discussion climate impinges upon the possibility of agonistic pluralism. Examples for (the few) instances where counterspeech acts did align with the strategies and values advocated by RI include efforts of users to point out and raise awareness about specific manipulative practices of digital hate culture at work in the comment section, posting links to information sources that undermine conspiracy theories and cautioning other counter speakers against lowering themselves to the level of those spreading hate (Figure 5).

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 5. Examples of counterspeech complying with RI standards



Note. Translated and visually reconstructed user comments.

Nevertheless, calls for a more objective and fact-based discussion and the refutation of disinformation and prejudices do not, in the majority of cases, show any effect on the discussion opponent, who typically refuses to reconsider her or his standpoint based on the explanation that the counter-argument stems from the "brainwashed leftist project" that has yet to be red-pilled (i.e., realize the truth). Counterspeech thus far shows limited potential for leading to consensus or stopping users from spreading hate; nevertheless, such thoughtful and fact- and conversation-oriented counter efforts do play a critical role in balancing the perceived public opinion. If the Islamophobic and hateful content was ignored, abiding by the motto "do not feed the troll," silent audiences would be left to assume that the articulated representations mirror the majority opinion. If we accept Mouffe's (2005) premise 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, the discussability that this form of organized rule-guided counterspeech promotes is crucial. Moreover, counterspeech can encourage other users with a similar attitude to get involved (Miškolci et al., 2018): Our data shows that comments of fact-bound reasoning and expressions of compassion for the targeted group receive a large number of "likes" and affirmative comments. The crux for assessing the potential of counterspeech to

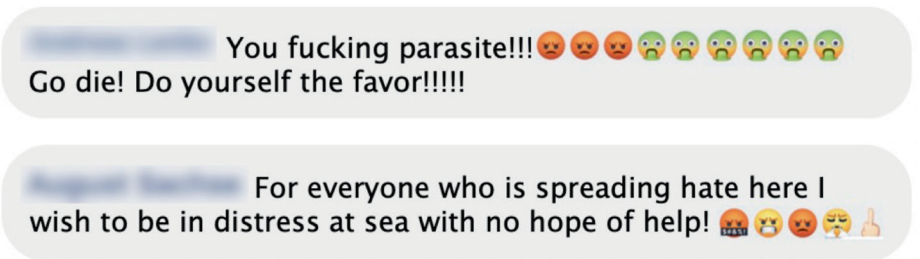
https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

transform antagonism into productive agonism thus remains to ask *how* counter acts are carried out.

5.2.2 Polarized camps of "racist neo-Nazis" and "deluded social justice warriors"

The large majority of user comments display a rigor and animosity with which the opposing sides fight over the truth that foster an antagonism separating society into a mutually incompatible left and right blocks. We identified those strategies and tactics promoted in RG's *Handbook for Media Guerillas* in comments addressing the respective opponent on *both* sides. The essence of trolling is obvious, and the repertoire of insults thrown at the discussion opponents wide. While one side calls the other "social justice warriors" (*Gutmenschen*), "color-naïve" (*Buntnaive*), and "leftist" (*Linksversiffte*), the other side replies with terms such as "racist," "right-wing extremist," "neo-Nazi," and "blue-brown" (*Blaubraune*), referring to the colors associated with the AfD and Third Reich National Socialism. Occasionally, threats and violent references occur, as demonstrated in this example from the data set (Figure 6).

Figure 6. User comments from both sides making use of abusive language

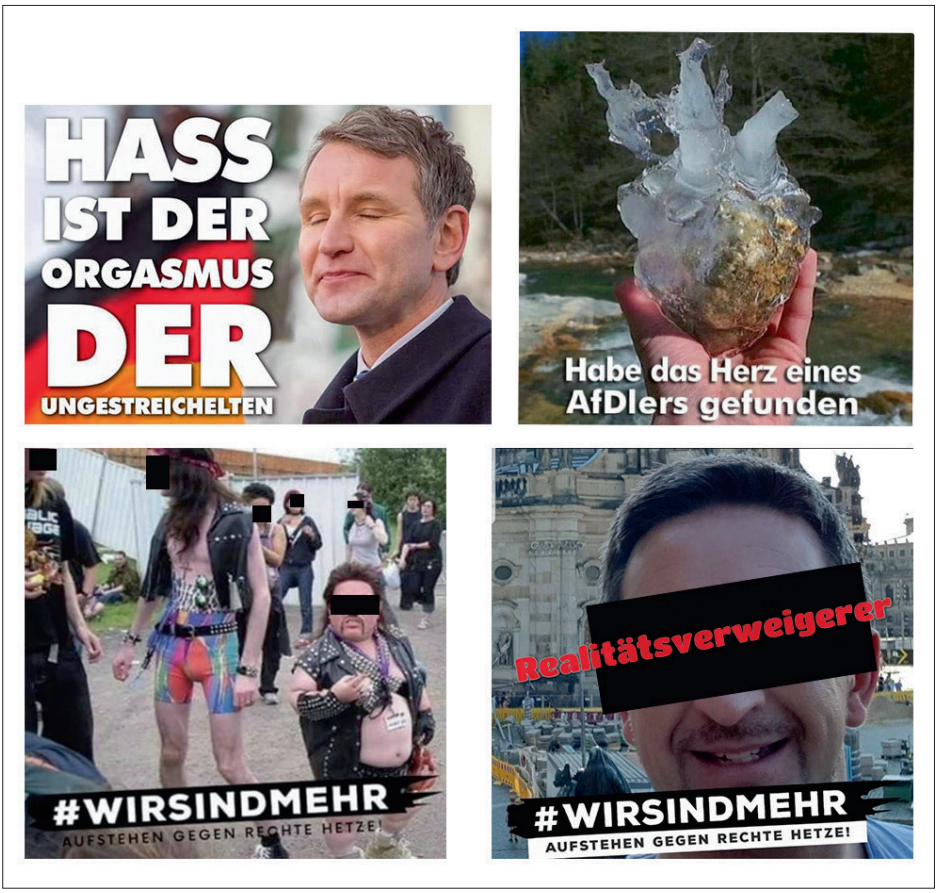


Note. Translated and visually reconstructed user comments from a user spreading hate towards those supporting sea rescue operations (above) and one user defending sea rescue operations (below).

The frequent appearance of memes and GIFs, appropriated to a German audience, show that humor and light entertainment are inseparable parts of political discussions online. Memes are used by both camps to display their mutual despise in a humoristic way. Contrary to Milner's (2013) findings, our analysis indicates that the "logic of lulz," although providing counter speakers with a powerful tool to appropriate Internet culture and appeal to (young) audiences in ways similar to hate groups, fails to be a productive means of encouraging a vibrant agonistic discussion in an adversarial sense. This example from our data (Figure 7) shows that it instead works to further antagonize core identity categories and essentialize binary dimensions of left and right. While the two first memes attempt to counter hate speech by mocking the right-wing populist party AfD, the ones below ridicule the "left camp" and label them as "reality deniers" by appropriating the anti-racist hashtag #wirsindmehr (#wearemore).

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 7. The antagonizing nature of the "logic of lulz" in a set of memes



Note. Translation top left: "Hate is the orgasm of the non-caressed" (picture: Björn Höcke, AFD politician). Top right: "I found the heart of an AfD supporter." Bottom left: "#wearemore – stand up against right-wing hatred!" Bottom right: "Reality denier. #wearemore – stand up against right-wing hatred!" (subsequently anonymized).

Again, what looms in this antagonistic discourse of left and right is the strong moralizing tone that Mouffe (2005) cautions against. Both sides seize every opportunity to discredit their opponent as evil, either by labeling users as "part of the deceitful system" or as right-wing extremists. This meme (Figure 8) illustrates how moralizing references about users' affinity to right-wing ideology and the use of the Nazi label, which are highly present in the analyzed discussions, ultimately risks playing into the hands of those spreading hate.

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Figure 8. A meme taking advantage of the excessive use of the Nazi label

Note. Translation: "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!"

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Rule-guided counterspeech, as promoted by RI, explicitly cautions against practices of othering and framing opponents as a moral enemy by using labels that assign political or ideological stances. Our analysis makes apparent how most users engaging in counterspeech lack (or ignore) this knowledge. Thus, it is crucial to promote a widespread understanding that taking too extreme positions and not complying with the civic and moral standards of organized counterspeech is counterproductive for the cause of efficiently refuting hate speech online.

6. Discussion and concluding reflections

This study demonstrates the need to develop a nuanced understanding of counterspeech and its different forms of expression to be able to adequately assess its (unintended) perils and the opportunities it holds to effectively refute online hate speech. While the intentions of organized initiatives like RI give the impression that counterspeech that adheres to civic and moral standards carries a promising agonistic potential, the counter efforts that dominate online discussions reflect a different reality. Despite being allegedly well-meaning, the majority of the counterspeech analyzed for this study only spurred on the polarization of discourse through what seem like spontaneous, imprudent reactions that leave little space for productive exchange. The investigation of commenting practices of users disseminating hate speech and those counter speaking, as well as the discourses that are embedded in their tug-of-war, showcases how popular social media today function as battlefields for partisan conflict while democratic politics fail to provide "a vibrant 'agonistic' public sphere of contestation where different hegemonic political projects can be confronted" (Mouffe, 2005, p. 3). The discursive antagonisms flourishing in the comment sections of Facebook are highly moralized, promoting a post-political we/they discrimination between good and evil in which the opponent is not perceived as a legitimate adversary but as an enemy to be destroyed. Users on both sides were more concerned with manifesting their own standpoint through mostly provocative and often offensive and abusive contributions instead of seeking productive exchange and mutual orientation. While acts of hate speech promoted hostility and fear towards Muslims as cultural and religious others and sought to convey an anti-establishment discourse that unveils the deceit of the "leftist system" against the "betrayed people," the large majority of the counterspeech efforts failed to adhere to the civic and moral standards of rule-guided counterspeech and, by stepping into similar moralizing and subjective communication practices as their opponents, only further spurred on the destructive dynamics.

6.1 Platformed hatred

For a holistic understanding of the nature of such antagonistic battles, we need to take into account the two remaining dimensions of Uldam and Kaun's (2019) framework for studying civic engagement online. Platform affordances and power relations essentially determine what can be said, how, by whom, and on which pages. The general rules for this might, in principle, be established by national

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

law (such as NetzDG), but the enforcement of such remains with private corporations. Although Facebook itself has shown recent efforts to extend its existing bans on hate speech (now prohibiting white nationalist content), studies suggest that despite all regulations, its technological affordances and platform policies are prone to serve the exclusionary antagonism that evolves in it (Matamoros-Fernández, 2017). Users who are familiar with the workings of Facebook's algorithms can deliberately manipulate them to serve their respective agendas (e.g., by maintaining fake profiles and social bots or by "upvoting" favorable content). While both organized hate groups and counter initiatives can benefit from algorithmic politics, the practicality of the manipulative practices of digital hate culture essentially *relies* on the exploitation of such – considering, for example, RG's advocated practices of "information wars," "social networking raids," and "memetic warfare." Research has shown that the execution of such coordinated attacks helps troll armies to trend on social media platforms and, in this way, dominate online discourse (Kreißel, Ebner, Urban, & Guhl, 2018). Our analysis confirmed that exponents of digital hate culture are well aware of respective platform policies and legal boundaries that need to be circumvented. As the *Handbook for Media Guerillas* reads, "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" (HoGeSatzbau, 2018). Further, although Facebook's community standards prohibit "[d]ehumanizing speech such as reference or comparison to … [s]exual predator, [s]ubhumanity, [v]iolent and sexual criminals" (Facebook, 2019), the prevalence of such in our data indicates that the tracking and deletion of content are not efficient enough. High accuracy, fully automated hate speech recognition is currently intractable (Schieb & Preuss, 2016); therefore, the decoding of hateful, sometimes humorous or ambiguous, statements comes down to Facebook's employees, whose success highly depends on sufficient human resources and the necessary contextual knowledge. The power of and responsibility for content moderation is further shared with the page hosts, in this case the news outlets, who can interfere in discussions or delete inappropriate comments. However, such measures cannot be regarded as a comprehensive solution; it remains an individual decision of the respective page host to act, and while some strive to ensure fair and respectful commenting, others might regard heated discussions as favorable because they increase the number of active users on their page.[8]

6.2 A call for more organized counteraction

Although social media platforms do not offer a promising agonistic perspective (based on the above), Facebook and co. will continue to be the channels where partisan conflict finds expression. Accepting the current condition and its disad-

8 This implies certain limitations for the methodological approach taken in this study: There can be no certainty as to how the different news outlets handled the deletion of hateful content before the comments were retrieved, which may have impacted the final sample in terms of actual versus "filtered" hate speech.

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access -

vantages, we argue that the antagonistic debates on social media can still be moved towards productive agonism. The practices of organized counterspeech, as advocated by civil society initiatives, carry the potential to foster agonistic pluralism in Mouffe's sense: With thorough knowledge about the opponent's manipulative practices and the appropriation of Internet culture, an initiative such as RI can provide a strong opposition to digital hate culture – one that is political, not moral, since it advocates seeing the opponent as an adversary who has a legitimate political standpoint and seeks mutual understanding through "love and reason." While such initiatives need to become stronger, it is still crucial to make those wishing to actively combat the workings of online hate aware that counterspeech is not a universally effective means and carries dangers when used thoughtlessly in the heat of the moment.

However, the question put forth by Cammaerts (2009) and Davids (2018) remains as to what limits tolerant agonism can offer a plausible response to hate speech. Since a productive agonistic discussion requires a minimal consensus on normative issues, including respect for human dignity, the often-dehumanizing practices of digital hate culture that became visible in the promoted anti-Muslim discourse certainly exceed these limits. Following Mouffe's argument, the exclusion of voices from public discourse and labeling them as "extreme-right" or "Nazi," would only exacerbate the problem. Nevertheless, 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 constitute a valuable and necessary means to combat cyberhate. Not only do their practices strive to correspond to the requirements of productive agonism, but their efforts also speak to publics beyond those actively participating in online debates: Well aware of the crucial role the perception of online debates plays for public discourse, organized counter initiatives spread an important message – namely that hate does not dominate public opinion and that it no longer goes unchallenged.

6.3 Future research

The insights gained from this study indicate several directions for future research that strives to undermine digital hate culture and advance civil society-led counter initiatives. First, the workings of organized online hate on popular social media platforms should be investigated further. Although the hubs of toxic discourses lay elsewhere on Alt-tech platforms and in the corners of the dark web, mainstream online platforms are 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 and investigating the potential effects of user comments on how people perceive the related online content and expressed opinions are fruitful paths for conducting such studies. For instance, in Germany, 41 percent of Internet users read user comments at least once a week, and 30 percent actively comment at least once a month (Ziegele, Jost, Bormann, & Heinbach, 2018). A thorough theoretical reflection and methodological operationalization of studying the role of platform affordances and

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

power relations in this regard would bring valuable additional insights and adds a holistic approach to the research phenomenon. Since our study was based on two concrete case studies, its generalizability is limited. Future research efforts, both on the workings of organized hate groups and on the potential of counter initiatives, are needed to support and complement our findings. These efforts should include further case studies, from a global context, that unpack the strategies of such groups and try to trace their digital footprints in the public spheres of social media. Another, and perhaps even more insightful, approach would be to investigate this phenomenon from an audience perspective. To better understand what counter efforts are able to unfold and how critically silent readers of such comment threads reflect on hate speech and manipulative practices online, it is necessary to include audiences in future research efforts.

When it comes to the future of civil-led counter initiatives, this study has shown that organized, communication- and exchange-oriented, rule-guided efforts can be effective to combat cyberhate. As such, they should be widely promoted and supported by more research that confirms their potential and helps to improve their activities. Current recommendations include that counterspeech needs to become even more dynamic, innovative, and bold (Ebner, 2019). However advanced, counterspeech initiatives cannot remain the only efficient response to the unrestrained hatred and poisoning of public discourse. To successfully fight and prevent digital hate culture, civil society actors, social media companies, and national legislators have to join forces. Building the basis for these response systems requires extensive research, without which it will not be possible to keep up with the rapidly evolving media ecosystems undergirding digital hate cultures, their internal dynamics, and their harmful influence on contemporary democratic societies.

References

Anders, R. (2017). *Lösch Dich! So organisiert ist der Hate im Netz* [Delete yourself! This is how organized online hate is]. Retrieved April 6, 2019 from https://www.youtube.com/watch?v=zvKjfWSPI7s

Askanius, T. (2021). On frogs, monkeys and execution memes: Exploring the humour-hate nexus at the intersection of neo-Nazi and Alt-right movements in Sweden. *Television & New Media* [Special issue: Extremism, Nationalism and Racism on Digital Platforms. Guest editors: Pantti, M., Nikunen, K. and Titley, G.], *22*(2).

Awan, I. (2016). Islamophobia on social media: A qualitative analysis of the Facebook's walls of hate. *International Journal of Cyber Criminology, 10*(1), 1–20. https://doi.org/10.5281/zenodo.58517

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), 362–380. https://doi.org/10.1093/bjc/azv122

Ayyadi, K. (2018). *Wie „Reconquista Germanica" auf Discord seine "Troll-Armee" organisiert* [How "Reconquista Germanica" organizes its "troll army" on Discord]. Retrieved April 6, 2019 from https://www.belltower.news/wie-reconquista-germanica-auf-discord-seine-troll-armee-organisiert-47020/

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access -

Benesch, S., Buerger, C., Glavinic, T., & Manion, S. (2018). *Dangerous speech: A practical guide*. Retrieved March 2, 2020 from https://dangerousspeech.org/guide/

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001

Blaya, C. (2018). Cyberhate: A review and content analysis of intervention strategies. *Aggression and Violent Behavior*, (45), 163–172. https://doi.org/10.1016/j.avb.2018.05.006

Boberg, S., Schatto-Eckrodt, T., Frischlich, L., & Quandt, T. (2018). The moral gatekeeper? Moderation and deletion of user-generated content in a leading news forum. *Media and Communication*, 6(4), 58–69. https://doi.org/10.17645/ mac.v6i4.1493

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* (pp. 137–153). Bielefeld: transcript. https://doi.org/10.14361/9783839446706-010

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), 555–575. https://doi.org/10.1177/1367877909342479

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* (pp. 155–167). Bielefeld: transcript. https://doi.org/10.14361/9783839446706-011

Davey, J., & Ebner, J. (2017). *The fringe insurgency: Connectivity, convergence and mainstreaming of the extreme right*. London: ISD. Retrieved May 25, 2019 from https://www.isdglobal.org/wp-content/uploads/2017/10/The-Fringe-Insurgency-221017.pdf

Davids, N. (2018). On the (in)tolerance of hate speech: Does it have legitimacy in a democracy? *Ethics and Education, 13*(3), 296–308. https://doi.org/10.1080/17449642.2018.1477036

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* (pp. 169–181). Bielefeld: transcript. https://doi.org/10.14361/9783839446706-012

Facebook (2019). *Community standards*. Retrieved May 15, 2019 from https://www.facebook.com/communitystandards/

Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet, 7*(3), 319–339. https://doi.org/10.1002/poi3.95

Ganesh, B. (2018). The ungovernability of digital hate culture. *Journal of International Affairs, 71*(2), 30–49. Retrieved January 15, 2019 from https://jia.sipa.columbia.edu/(...) digital-hate-culture

George, C. (2015). Hate speech law and policy. In R. Mansell, & P. H. Ang (Eds.), *The international encyclopedia of digital communication and society* (pp. 305–314). Malden, MA: Wiley Blackwell. https://doi.org/10.1002/9781118767771.wbiedcs139

George, C. (2016). Regulating "hate spin": The limits of law in managing religious incitement and offense. *International Journal of Communication, 10*, 2955–2972. Retrieved April 6, 2019 from https://repository.hkbu.edu.hk/hkbu_staff_publication/6440

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Gollatz, K., Riedl, M. J., & Pohlmann, J. (2018). *Removals of online hate speech in numbers*. Retrieved April 6, 2019 from https://zenodo.org/record/1342325#.XKsOSy-Bqu4

Guhl, J., Ebner, J. and Rau, J. (2020). The online ecosystem of the German far-right. London: ISD, Institute for Strategic Dialogue. Retrieved March 25, 2020 from https://www.bosch-stiftung.de/sites/default/files/publications/pdf/2020-02/ISD-The%20Online%20Ecosystem%20of%20the%20German%20Far-Right.pdf

HoGeSatzbau (2018, January). *D-Generation*. Retrieved May 25, 2019 from https://www.hogesatzbau.de/wp-content/uploads/2018/01/HANDBUCH-FÜR-MEDIENGUERILLAS.pdf

Jaki, S., & De Smedt, T. (2018, May). *Right-wing German hate speech on twitter: Analysis and automatic detection*. Retrieved April 6, 2019 from http://www.organisms.be/downloads/jaki2018.pdf

Kreißel, P., Ebner, J., Urban, A., & Guhl, J. (2018). *Hass auf Knopfdruck: Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz* [Hate at the touch of a button: Right-wing extremist troll factories and the ecosystem of coordinated hate campaigns online]. London: ISD. Retrieved 20 February, 2019 from https://www.isdglobal.org/wp-content/uploads/2018/07/ISD_Ich_Bin_Hier_2.pdf

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage. https://doi.org/10.2307/2288384

Landesanstalt für Medien NRW (2018). *Ergebnisbericht Hassrede* [Results report hate speech]. Düsseldorf: forsa. Retrieved 15 January, 2019 from https://www.medienanstalt-nrw.de/zum-nachlesen/forschung/abgeschlossene-projekte/forsa-befragung-zur-wahrnehmung-von-hassrede.html

Lauer, S. (2018). *Infokrieg für die Grundschule: „Reconquista Germanica" und „D-Generation"* [Information war for primary school: "Reconquista Germanica" and "D-Generation"]. Retrieved April 6, 2019 from https://www.belltower.news/infokrieg-fuer-die-grundschule-reconquista-germanica-und-d-generation-46716/

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), https://doi.org/*555–579*. 10.*5771/2192-4007-2018-4-555*

Lepoutre, M. (2017). Hate speech in public discourse: A pessimistic defense of counterspeech. *Social Theory and Practice, 43*(4), 851–883. https://doi.org/10.5840/soctheorpract201711125

Lück, J., & Nardi, C. (2019). Incivility in user comments on online news articles: Investigating the role of opinion dissonance for the effects of incivility on attitudes, emotions and the willingness to participate. *Studies in Communication and Media, 8*(3), 311–337. https://doi.org/10.5771/2192-4007-2019-3-311

Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. New York: Data & Society Research Institute. Retrieved April 6, 2019 from https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf

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), 930–946. https://doi.org/10.1080/1369118X.2017.1293130

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution.* Retrieved April 6, 2019 from https://www.ssoar.info/ssoar/handle/document/39517

Milner, R. M. (2013). FCJ-156 hacking the social: Internet memes, identity antagonism, and the logic of lulz. *The Fibreculture Journal, 22*, 62–92. Retrieved January 13, 2019 from http://twentytwo.fibreculturejournal.org/fcj-156-hacking-the-social-internet-memes-identity-antagonism-and-the-logic-of-lulz/

Miškolci, J., Kováčová, L., & Rigová, E. (2018). Countering hate speech on Facebook: The case of the Roma minority in Slovakia. *Social Science Computer Review*, 1–19. https://doi.org/10.1177/0894439318791786

Mouffe, C. (2005). *On the political.* London: Routledge. https://doi.org/10.1057/palgrave.cpt.9300293

Mouffe, C. (2016). Democratic politics and conflict: An agonistic approach. *Política común, 9*. Retrieved January 10, 2020 from https://quod.lib.umich.edu/p/pc/12322227.0009.011?view=text;rgn=main

Müller, K., & Schwarz, C. (2017). *Fanning the flames of hate: Social media and hate crime.* Retrieved April 6, 2019 from https://doi.org/10.2139/ssrn.3082972

Phillips, W. (2018). *The oxygen of amplification: Better practices for reporting on extremists, antagonists, and manipulators online.* New York: Data & Society Research Institute. Retrieved March 25, 2019 from https://datasociety.net/library/oxygen-of-amplification/

Reconquista Internet (2018, 25 May). *Reconquista Internet – mit Herz und Hand und mit Verstand* [Reconquista Internet – with heart and hand and with reason]. Retrieved April 6, 2019 from https://www.facebook.com/pg/ReconquistaNetz/about/?ref=page_internal

Reconquista Internet (2019). *Reddit*. Retrieved April 6, 2019 from https://www.reddit.com/r/ReconquistaInternet/wiki/index#wiki_reconquistainternet_wiki

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), 459–472. https://doi.org/10.5771/2192-4007-2018-4-459

Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics, 22*, 463–487. https://doi.org/10.1177/1940161211415849

Schieb, C. & Preuss, M. (2016). *Governing hate speech by means of counter speech on Facebook*. In 66th ICA Annual Conference at Fukuoka, Japan. Retrieved April 6, 2019 from https://www.researchgate.net/publication/303497937_Governing_hate_speech_by_mea ns_of_counterspeech_on_Facebook

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), 580–606. https://doi.org/10.5771/2192-4007-2018-4-580

Schwarzenegger, C., & Wagner, A. J. (2018). Can it be hate, if it is fun? Discursive ensembles of hatred and laughter in extreme right satire on Facebook. *Studies in Communication and Media, 7*(4), 473–498. https://doi.org/10.5771/2192-4007-2018-4-473

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.

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND

Retrieved May 15, 2019 from https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf

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), 502–522. https://doi.org/10.1080/13602004.2016.1259054

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* (pp. 193–204). Bielefeld: transcript. https://doi.org/10.14361/9783839446706-014

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* (pp. 101–115). London & New York: Routledge. https://doi.org/10.4324/9781315107066-8

Zick, A., Krause, D., Berghan, W., & Küpper, B. (2016). Gruppenbezogene Menschenfeindlichkeit in Deutschland 2002–2016. [Group-based enmity in Germany 2002–2016.] In A. Zick, B. Küpper, & D. Krause (Eds.), *Gespaltene Mitte – Feindselige Zustände: Rechtsextreme Einstellungen in Deutschland 2016* [Split center – hostile conditions: Far-right attitudes in Germany 2016] (pp. 33–81). Bonn: Dietz.

Ziegele, M., Jost, P., Bormann, M., & Heinbach, D. (2018). Journalistic counter-voices in comment sections: Patterns, determinants, and potential consequences of interactive moderation of uncivil user comments. *Studies in Communication and Media, 7*(4), 525–554. https://doi.org/10.5771/2192-4007-2018-4-525

https://doi.org/10.5771/2192-4007-2020-4-540 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-NC-ND