

Who is responsible for interventions against problematic comments? Comparing user attitudes in Germany and the United States

Martin J. Riedl¹  | Teresa K. Naab²  | Gina M. Masullo¹  |
Pablo Jost³  | Marc Ziegele⁴ 

¹School of Journalism, The University of Texas at Austin, Austin, Texas, USA

²Department of Media, Knowledge and Communication, University of Augsburg, Augsburg, Germany

³Department of Communication, Johannes Gutenberg-University Mainz, Mainz, Germany

⁴Department of Social Sciences, Heinrich Heine University Dusseldorf, Düsseldorf, Germany

Correspondence

Teresa K. Naab, Department of Media, Knowledge and Communication, University of Augsburg, Universitaetsstrasse 10, 86159 Augsburg, Germany.
Email: teresa.naab@phil.uni-augsburg.de

Funding information

Center for Media Convergence, Johannes Gutenberg University of Mainz; William and Flora Hewlett Foundation

Abstract

Online comment sections on news organizations' social media pages provide a unique forum for exploring attitudes toward platform governance and freedom of expression at the crossroads between people, platforms, and news providers. Amid ample political and policy interest, little empirical evidence exists on user perceptions of platform governance. Through survey studies in Germany ($n = 1155$) and the United States ($n = 1164$), we provide a comparative perspective on responsibility attributions toward different regulatory actors who may intervene against problematic user comments: the state (law enforcement), platform operators (Facebook), news organizations, and users themselves. We explore this against the backdrop of different notions of free speech and cultural differences in the two countries. We find that Germans attribute greater responsibility for intervention to the state, Facebook, and news organizations than Americans. They also assume greater self-responsibility. While support for free speech did not impact responsibility attribution to Facebook, news organizations, or the users themselves, people with greater general support for free speech saw law enforcement as less responsible for intervention. The results provide empirical evidence for an integrated view of various regulatory actors who complement each other in the governance of online discussions.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Policy & Internet* published by Wiley Periodicals LLC on behalf of Policy Studies Organization

KEYWORDS

content moderation, Facebook, free speech, Germany, Network Enforcement Act, online comments, online intervention, regulation, United States

INTRODUCTION

Online comment sections on the social media pages of news outlets provide spaces for readers to express their opinions and can constitute a forum for less-represented views (Dahlberg, 2011). Particularly, comment sections on social media pages with great reach such as on Facebook are important fora of public debate (Esau et al., 2017). But these online discussions also include the voices of the intolerant and hateful. A relevant share of comments includes incivility in the forms of name-calling, aspersion, lying, vulgarity, pejorative speech, racism, and sexism (Chen, 2017; Coe et al., 2014). Empirical research has documented various negative effects of incivility in comment sections on readers and society (e.g., Anderson et al., 2014; Chen, 2017; Ziegele et al., 2018). Users, too, perceive vitriolic comments as problematic and in need of intervention (Duggan, 2017). As a consequence, societies must find ways to regulate such spaces to preserve the potential benefits of comment sections for public debate and simultaneously protect minority voices.

This paper responds to calls for more survey-based research on users' perspectives on regulation and intervention (Einwiller & Kim, 2020). Ample political, policy, and research attention has been paid to the legal frameworks that govern liability regimes for problematic user-generated content, particularly as these differ between countries and jurisdictions. Additionally, many practitioners and scholars deal with the ways in which platforms and news organizations handle moderation of this content (Daskal et al., 2020; Klonick, 2017). This, too, varies between countries.

However, little empirical evidence is available on what users think about interventions by these actors. Presumably, perceptions will reflect differences between countries. Public opinion data can stimulate debates about existing intervention strategies of various actors. In line with the idea of evidence-based policymaking (Livingstone, 2013; Puppis & Van den Bulck, 2019), such research can inform policymakers about the public's understanding of media regulation, identify deviations between policy and public opinion, and help them to justify or adjust their decisions (Lambe, 2002; Naab, 2016; Naab et al., 2019; Paek et al., 2008). As Internet governance regimes consider the role that users themselves might play in regulation, it is crucial to provide independent academic assessments of their attitudes. Research such as ours can also contribute to platforms taming their reluctance toward regulation if research shows user perceptions may be in favor of regulation. Studies on public perceptions of free speech and regulation can also serve educational programs and the long-term protection of free speech. A comparative view might be able to generate a deeper understanding of different national approaches. This seems particularly relevant as online content is distributed and received transnationally, yet regulation mostly rests on national actors.

This paper makes a novel contribution to our understanding of who the public thinks should assume responsibility for interventions against problematic comments in a comparative framework between the United States and Germany. Comparative research on user perceptions on regulation is theoretically important as it provides a better understanding of different regulatory approaches to content governance. It also allows for an integrated view on different actors of Internet governance who complement each other in the regulation of online discussions. We acknowledge distinct cultural and country-specific factors that impact

freedom of expression and policy regulating technologies related to speech (Choi et al., 2012). Most importantly, we consider that users' notions of social media regulation will strongly depend on their support for freedom of expression, along with other factors.

We engage in a comparative study in Germany ($n=1155$) and the United States ($n=1164$) to investigate peoples' perceptions of who should be responsible for interventions against problematic comments on the Facebook pages of news outlets. Germany and the United States are important cases: Germany is defined by a culture of freedom of expression firmly rooted in human dignity and the possibility of banning certain speech that counters constitutional principles. In 2017, Germany even introduced rigid national legislation to create liability regimes for social media platforms through the Network Enforcement Act (Heldt, 2019). The US conception of freedom of expression, on the contrary, is firmly rooted in liberty and an almost religious reverence for the First Amendment. In the US, technology regulation striving for more liability for social media platforms is similarly gathering steam, pointing to the importance of considering attitudes toward regulation in conjunction with attitudes toward free expression. Together, these two Western democracies with their distinct approaches to freedom of expression provide a productive backdrop to discuss user perceptions of who should be responsible to intervene in online news comment sections on social media.

ACTORS INTERVENING IN COMMENT SECTIONS

We conceive of intervention as an umbrella term for a variety of activities related to dealing with problematic user-generated content, including deciding whether content should be published, removing content that violates policies, and engaging with authors who deviate from perceived standards (Masullo et al., 2020; Riedl et al., 2020). Various actors may be involved in interventions against problematic speech in news comment sections. These actors differ in the specific interventions that they might conduct, their license to do so, and their efficacy. Regulation of traditional media content as well as speech was and still is codified and executed by nation-states and state-like actors, such as the European Union (Balkin, 2014). However, the emergence of technology companies and free license for users to produce content has given way to a “pluralist model of speech regulation” (Balkin, 2017, p. 1189). Regulation must now happen in a multi-stakeholder fashion, wherein speech is also governed by speakers themselves (e.g., citizens, news media, or other private corporations). Under the term *intervention* we subsume these various activities of dealing with problematic content—different regulatory actors and their grasp on regulation. We differentiate the intervention strategies of these actors as follows:

The state as a regulatory entity legislates, maintains, and enforces laws through law enforcement authorities. Laws created by elected representatives of the state govern relationships between users, technology platforms, and state power. They qualify when and how law enforcement may get involved or when platforms do not conduct themselves adequately.

Platform operators, such as Facebook, host social media pages of news outlets, and set up governance regimes of acceptable content that emerge within the perimeters of existing laws and their own transnational and commercial interests. Their regulation is contingent on trust and safety teams developing codes of conduct and community guidelines that enshrine norms and rules of user engagement on platforms. These are then enacted and monitored by algorithms, filters, human moderators, and combinations thereof (Myers West, 2018; Roberts, 2019).

News organizations provide journalistic content that is then commented upon in the news organizations' comment sections. Many organizations moderate uncivil comments by

filtering or deleting them manually or with the help of automated tools (Stroud et al., 2015). Furthermore, media organizations consider ways in which professional community moderators could respond to comments (Naab et al., 2020; Wright et al., 2019; Ziegele & Jost, 2020). Generally, comment moderation by media organizations is assumed to be a crucial element for successful online discussions (Kraut & Resnick, 2012; Wright, 2006), and empirical research backs up that moderation, together with other policies, indeed benefits civility (Ksiazek, 2015). Users, too, indicate a higher willingness to participate in moderated over unmoderated online discussions (Wise et al., 2006).

Lastly, *users* can intervene against content perceived as problematic. Volunteer moderation by users exists in various spaces, and researchers have referred to it as a form of civic labor in the digital space (Matias, 2019), online civic intervention (Porten-Chée et al., 2020), or a specific type of corrective action (Lim, 2017; Naab et al., 2019; Ziegele et al., 2019). Most social media platforms, including Facebook, provide the opportunity to flag the content of others that potentially violates the commenting policy of the discussion provider. Users can also choose to downvote comments they perceive as inappropriate or intervene against comments of others through counter-speech comments (Kalch & Naab, 2017; Kim, 2021; Naab et al., 2018, 2019). While in traditional media settings users were dependent on institutions to take action against problematic content, social media empower ordinary users to engage against content they perceive as problematic (Lim & Golan, 2011).

ATTRIBUTING RESPONSIBILITY FOR INTERVENTION

The emergence of various actors with opportunities to intervene against content brings to the fore what is referred to as the “problem of many hands” (Thompson, 2014, p. 259). It describes the difficulty of attributing responsibility for specific decisions in complex processes in which many parties are involved, spurring a need for “cooperative responsibility” (Helberger et al., 2018, p. 1). Intervening against problematic comments hosted on the pages of news outlets on social media may fall into shared responsibilities between all four groups of actors that we outlined above—the state, platform operators, news organizations, and users. The important question to ask is how this matches against user perceptions. Consider, for example, if users primarily attribute responsibility to institutional actors, the success of counter-speech may be moot. Furthermore, users' ideas of who should intervene could influence the public discourse on platform regulation, as well as show platforms what their users want.

Previous literature has considered the attitudes of citizens (or more specifically of media users) toward interventions against media content (Lambe, 2002). Empirical studies have focused on users' approval of state interventions, for example, whether certain media content should be censored, if communicators should be punished, or if content should be assessed before publication (e.g., Golan & Lim, 2016; Lambe, 2002; Lim, 2017). Some studies have asked if users approve of restrictive actions by media companies or Internet providers or if they demand actions by these actors (Lo & Wei, 2005; Wei & Lo, 2007). Rarely has research looked into whether users approve that other users intervene against content (Wei & Lo, 2007) or examined self-engagement against certain content (Lo & Wei, 2005; Rojas et al., 1996). Self-responsibility is the “accountability of an individual or collective actor for actions that have already been performed or are going to be performed in the future” (Maier, 2019, p. 27). In comment sections, perceived self-responsibility refers to the question of whether people perceive that they themselves have a duty to engage in interventions against uncivil online comments.

As research is typically interested in examining public approval of one specific actor and their intervention(s), it seldomly allows for comparisons of perceived responsibility

attributions between different actors. Our paper addresses this void and examines which actors users perceive should intervene.

INFLUENCES OF NATIONAL BACKGROUND ON RESPONSIBILITY ATTRIBUTIONS

As we investigate individual responsibility attributions to various actors, we argue that these attributions may be contingent on users' concepts of freedom of expression and their support for free speech in general. People with disparate national backgrounds as well as people with different appreciation for free speech might differ in their perceptions of who should be responsible to intervene against uncivil comments.

Attributing responsibility to the state

Germany and the United States, as most Western democracies, set strong constitutional limits to potential state interference in media content including social media content like online comments. However, differences emerge when thinking about the possibility of restricting freedom of expression (Kommers, 2019).

In the United States, free speech is guaranteed by the First Amendment to its Constitution, which forbids the government from banning speech, except in certain instances. This concept of freedom of expression is best summarized as free speech exceptionalism (Krotoszynski, 2015). It is tethered to individualism and independence—against the backdrop of possible government interference (Krotoszynski, 2015; Nieuwenhuis, 2000; Post, 2017). American free speech harkens to the notion of a marketplace of ideas, much in the sense of John Stuart Mill's articulations in *On Liberty* (Medeiros, 2017). In this purview, some types of speech—for example, racist speech or other forms of hate speech—have to be permitted (Nieuwenhuis, 2000), with the underlying idea that truth will emerge when ideas compete in a marketplace of ideas. While the US approach to free speech may seem far-reaching and idiosyncratic, it is also not a *carte blanche* for any type of speech in any setting and venue. Exceptions include speech that instigates breaking the law in a manner that is both imminent and likely (Horwitz, 1991) or defamation, to name but a few. Furthermore, support for the US view on freedom of expression remains in flux as society changes (Chong & Levy, 2018; Medeiros, 2019).

The German Constitution prohibits any kind of state control of media content aiming to censor before publication. Nevertheless, the German Constitution enshrines the notion of human dignity above all (Carmi, 2007; Hawdon et al., 2017). This may entail limiting the speech rights of some to protect the dignity of others (Krotoszynski, 2015). The German concept of free speech assumes that to protect democracy itself it may be necessary to forbid some forms of speech, namely speech that counters the very premises of the democratic system (Khan, 2013). Notably, whereas certain kinds of hate speech are prohibited in Germany (e.g., Holocaust denial), similar laws do not exist in the United States. But even though the German notion of free speech is more restrictive, restrictions to speech are still viewed with suspicion (Nieuwenhuis, 2000).

Regarding the attitudes of Germans and Americans toward state interventions against speech, Nieuwenhuis (2000) points to an important difference between the two peoples: Germans trust the state more than Americans (Wike et al., 2017). For example, in 2011, Pew Research Center's Global Attitudes Project (Kohut et al., 2011) published data measuring people's attitudes toward the role of state interference in society. While in the United States, 58% agreed that one should pursue life goals without state interference, in Germany,

it was only 36%. In particular, when asked about the importance of the media reporting the news without state censorship, 80% of Americans say that this is very important, but only 67% of Germans do so (Wike et al., 2019). Based on these country-specific differences in the concepts of free speech and in the attitudes toward state intervention, we hypothesize:

H1 *Germans will attribute greater responsibility for interventions against problematic online comments to law enforcement (the state) than Americans.*

Attributing responsibility to platform operators

Interventions by platform operators are based on fundamentally different approaches in Germany and the United States. The United States is known for its safe harbor protections for technology companies, enshrined in Section 230 of the Communications Decency Act of 1996 (Medeiros, 2017). This shields platforms from liability if their users post illegal content and also allows platforms to behave as good Samaritans (Citron & Wittes, 2017): they may police according to their own rules, while not being held accountable for errors or oversights. Meanwhile, research finds that 64% of Americans think online services (including social media platforms) should “play a major role” in tackling harassment (Duggan, 2017, p. 6). What unites the two countries is the important role that Facebook plays; in both Germany (22%) and the United States (35%) it is the most important social media site for news when compared with other social media platforms (Newman et al., 2020).

In Germany, in a marked difference to the US approach, operators of social media platforms are explicitly held accountable by the German Network Enforcement Act. Following this law, social media platforms with more than 2 million users in Germany must respond to content takedown requests for manifestly unlawful content within specific time frames (Heldt, 2019; Schulz, 2018). Otherwise, they might be fined up to 50 million Euros. The law is criticized by civil society organizations, certain political parties, and academics alike (Gollatz et al., 2018). They claim that its provisions could lead to “overblocking”—preemptively censoring speech that may still be within the limits of the law (Daskal et al., 2020). It is also criticized because the German state effectively outsources the policing of speech and its penal code to predominantly US-based private corporations. Despite this critique, survey research at the time found support among the German population for the draft law: 43% of Germans were definitely in favor of a law regulating hate speech and fake news. About a third of the population was also concerned about the ramifications such a law might have for freedom of expression (Inhoffen, 2017).

Considering that Germany had passed a law that assigns responsibility to social media operators, and that the German public is somewhat familiar with the law and surrounding discourses, one could assume that Germans would display greater support for interventions in comment sections by platforms than Americans would. We posit the following:

H2 *Germans will attribute greater responsibility for interventions against problematic online comments to Facebook than Americans.*

Attribution of responsibility to news organizations

While research has suggested that news organizations are important actors in the regulation of comment sections (Wright, 2006), little is known about cultural differences regarding users' perceptions of content moderation by news outlets. Germany and the United States lend themselves to a comparative perspective as their media systems are substantively different.

Germany has a strong public service broadcasting system, and state intervention—for example in the form of press subsidies—is the norm, whereas the US system touts the free market (Brüggemann et al., 2014). Evidence suggests further that Germans trust the news media more than Americans. Research on trust in the media finds that more than 40% of Germans indicate they trust the news, while less than a third of Americans do (Hanitzsch et al., 2018; Newman et al., 2020). Assuming that a general level of trust in the news media translates to trusting news organizations in their approach to intervening in comment sections, we posit:

H3 *Germans attribute greater responsibility for interventions against problematic online comments to news organizations than Americans.*

Assuming self-responsibility

In the United States, 60% of people think that bystanders who witness online harassment should play a major role in tackling harassment (Duggan, 2017). In Germany, survey research shows that approximately 20% of Internet users say they have engaged in some form of counterspeech against hate speech already (Landesanstalt für Medien NRW, 2020). Einwiller and Kim (2020) analyzed policy documents in which platforms across the United States, Germany, Korea, and China articulate their approaches toward harmful online communication. Among other dimensions, they find that platforms rarely encourage counterspeech explicitly. Perhaps the most prominent example of counterspeech—when it occurs—is the social online movement #ichbinhier (i.e., #iamhere; Ziegele et al., 2019). The Facebook group, a spin-off of the Swedish group #jagärhär, was founded in 2016 and currently has more than 45,000 members. The group strives to alleviate the quality of online news discourse in comment sections through collective action—by intervening in uncivil conversations with respectful counterspeech (Ley, 2018). To our knowledge, no similar initiative exists in the United States. This suggests that Americans compared with Germans might be less willing to assume self-responsibility to intervene in comment sections.

However, when comparing attitudes toward individualism and the capacity to bring about change, Americans assume more individual agency to change things than Germans. For example, asked whether people “disagree that success in life is pretty much determined by forces outside our control,” only 31% of Germans but as much as 57% of Americans said that they did (Wike, 2016). Indeed, internal political efficacy and social media self-efficacy beliefs are positively related to participation in social media (Velasquez & LaRose, 2015). Additionally, survey research on members of the group #iamhere suggests that perceived self-efficacy in writing counter-speech increases the members' likelihood to engage against problematic comments. However, general political efficacy had no effect (Ziegele et al., 2019). Given the lack of a clear direction from the literature, we pose a research question:

RQ1 *Is there a difference between Germans and Americans in assuming self-responsibility for interventions against problematic online comments?*

INDIVIDUAL SUPPORT FOR FREE SPEECH

Support for free speech also matters at the individual level: Some people are more able or willing to consider the value of free speech as a general good to individuals and society (e.g., World Values Survey, 2019). An individual's appreciation for free speech is a crucial indicator for the status of free speech as it gives information on its importance for those expected to exercise and advocate this right (Inglehart, 1997).

Various studies have highlighted that people who place greater value on free speech are more inclined to grant it even to undesirable or extreme expressions and show less approval of restrictions (Paek et al., 2008). However, research has not yet considered whether the individual level of appreciation of free speech also influences individuals' assessment of regulatory actors.

Influences on responsibility attribution for interventions

In light of scholarship insinuating that “[o]ne of the great evils free speech is meant to defend against is a state orthodoxy” (Corbin, 2009, p. 993), we consider how individual support for free speech relates to responsibility attributions to the state to intervene against problematic user comments: We expect people who display stronger support for free speech to hold more skeptical attitudes toward state interference in comment sections. This is based on the idea that any authoritative state restriction of free speech might be viewed as potential censorship and thus a harm to freedom of expression. Legally (in Germany and the United States), censorship occurs only when the state restricts content (Lambe, 2002).

People who strongly support free speech would also be expected to be suspicious not only about state interference but also about interventions by other institutionalized actors, such as Facebook, who hold a dominant position in society. We consider Facebook an institutionalized actor as it constitutes a formal organization that governs the individual behaviors of its staff and of users of the platform. Facebook presents its own terms of service including how inappropriate content is dealt with. Beyond this immediate context, the company has a tremendous impact on online ecosystems through its “algorithmic and data-driven practices” (Caplan & boyd, 2018, p. 2). Its corporate impact has led toward monopolization, to the extent that in December 2020, the US Federal Trade Commission launched an antitrust lawsuit against the company (Kang & Isaac, 2020). Research has shown that users who indicate greater commitment to free speech are also more skeptical toward restrictions that Facebook or law enforcement employ to limit uncivil online comments than those users who have lower commitment to free speech (Naab et al., 2019). Therefore, we postulate:

H4 *Greater individual support for free speech decreases attribution of responsibility for interventions against problematic comments (a) to law enforcement and (b) to Facebook.*

News media are often conceptualized as the Fourth Estate, observing and controlling other democratic institutions, and, thus, fulfilling an important societal function (Hampton, 2010). The media's very existence depends on free speech and protection from interference by the state and powerful groups. In this vein, the news media may be perceived as an appropriate actor to moderate citizens' expression without endangering free speech. On the contrary, the moderation strategies of news organizations have also been criticized for inadequately limiting the Internet's potential for citizen participation and free speech (Janssen & Kies, 2005; Wright, 2006). Additionally, both in Germany and the United States, a significant part of society perceives the news media as untrustworthy and part of a culprit elite (e.g., Fawzi, 2019). Given the lack of empirical evidence specifically regarding interventions against problematic comments, we ask:

RQ2 *Does support for free speech influence attribution of responsibility for interventions against problematic online comments to news organizations?*

Previous research did not find any impact of individual support for free speech on user interventions against uncivil comments (Naab, 2016; Naab et al., 2019). This is surprising, because one might expect that supporters of free speech, in particular, would rely on the self-regulatory power of engaging in discussions. Against this backdrop, we ask:

RQ3 *Does support for free speech influence whether people assume self-responsibility for interventions against problematic online comments?*

Relationship between national background and support for free speech

Data comparing support for free speech between Germans and Americans indicate that Germans, on average, show a slightly higher appreciation for free speech as a human right. For example, the Pew Research Center (Wike et al., 2019) found that 77% of Americans state that it is very important to have free speech in their country, compared with 86% of Germans. In line with this, data on people's individual evaluations of freedom of expression have been continuously gathered by the World Values Survey (2019): More than 20% of Germans compared with 17% of Americans chose "protecting freedom of speech" as their most important value. With regard to attitudes toward freedom on the Internet, research has found that support for freedom of expression on the Internet and on the necessity for governance to alleviate harm range at about the same level in Germany and the United States, at around 85% (Shen, 2017). Based on this data, we hypothesize:

H5 *Germans will display higher levels of general support for free speech than Americans.*

METHODS

Procedure and participants

The German and the US survey obtained Institutional Review Board approval on November 20, 2018. The German survey was launched in December 2018, and the US survey followed in January 2019. Procedures for both surveys were identical, but questions were worded in the German language for the German study and in English for the US study. We recruited US and German participants through Dynata,¹ an international survey company that provides researchers with access to panels that aim to represent the demographics of the adult population in each country. Table 1 presents the demographics of the US sample ($n = 1164$) and the German sample ($n = 1155$) and compares the final sample after data cleaning with the demographics of the Internet population in each country.

Measures

Attribution of responsibility for interventions against uncivil online comments on the Facebook pages of news outlets was measured with one item for each type of actor ("In comment sections on the Facebook pages of news media outlets, there are often comments that many people perceive as offensive, inappropriate, or problematic. In your opinion, who is responsible for doing something about such comments?"): "law enforcement," "Facebook," "the news organizations who maintain the respective Facebook pages," and "me myself."

TABLE 1 Demographics of participants compared with the German/US Internet populations

	German survey		US survey	
	Internet population (%)	Sample (%)	Internet population (%)	Sample (%)
Gender				
Male	50	48	50	41
Female	50	52	50	59
Race/ethnicity (US only)				
White	-	-	70	75
Black or African-American	-	-	13	10
Asian or Asian-American	-	-	5	3
Other	-	-	12	5
Hispanic (yes/no)	-	-	15	12
Age				
18–29	19	15	24	20
30–49	38	36	36	40
50–64	30	34	25	26
65+	13	15	15	15
Education				
High school or less (US), level 1 (GER)	29	21	17	4
Secondary (US), level 2 (GER)	34	31	17	20
Some college (US), level 3 (GER)	16	27	33	39
College + (US), level 4 (GER)	21	22	33	37
<i>N</i>		1155		1164

Agreement was measured on a scale from 1 = *strongly disagree* to 5 = *strongly agree*. Table 2 presents the frequencies by country. Table 3 presents the correlations between responsibility attribution to the four actors.

Support for free speech was measured with three items adapted from Rojas et al. (1996): “Everybody should have the freedom to publicly say what they believe to be true,” “No matter how controversial an idea is, an individual should be able to express it publicly,” and “All individuals should have the right to openly express their ideas, no matter how prejudiced they might be.” Agreement was measured on a scale from 1 = *strongly disagree* to 5 = *strongly agree* (German sample: $M = 4.05$, $SD = 0.85$, $\alpha = 0.88$; US sample: $M = 3.80$, $SD = 0.96$, $\alpha = 0.88$).

Country of the participants was coded 0 = Germany and 1 = United States.

Controls. Socio-demographic variables have been shown to be related to attitudes toward media regulation and free speech (Lambe, 2002). Moreover, news consumption, social media use as well as social media political self-efficacy are associated with individuals' social media engagement for their causes, their perception of harmful content and

TABLE 2 Descriptive statistics of attribution of responsibility to various actors by country

	Germany (<i>n</i> = 1155)		US (<i>n</i> = 1164)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Facebook	3.89	1.18	3.70	1.25
The news organizations who maintain the respective Facebook pages	3.79	1.14	3.58	1.25
Law enforcement	3.30	1.24	2.60	1.25
Me myself	2.94	1.37	2.81	1.38

TABLE 3 Pearson's correlations between attribution of responsibility to law enforcement, Facebook, news organizations, and self-responsibility

	Responsibility attribution to Facebook	Responsibility attribution to news organizations	Self-responsibility
Responsibility to law enforcement	0.410**	0.388**	0.285**
Responsibility to Facebook		0.588**	0.194**
Responsibility to news organizations			0.195**

Note: *N* = 2319.

p* < 0.05; *p* < 0.01; ****p* < 0.001.

intervention against such (Kalch & Naab, 2017; Kenski et al., 2017; Naab et al., 2018; Velasquez & LaRose, 2015; Ziegele et al., 2019). Thus, we measured gender, age, education (0 = secondary education or less; 1 = more), and political ideology (0 = liberal; 10 = conservative; German sample: *M* = 4.84, *SD* = 2.00; US sample: *M* = 5.06, *SD* = 3.10). Furthermore, we measured Facebook use for news and commenting on news with four items such as "I read content from news media outlets on Facebook"; 1 = *never*; 6 = *very often*; German sample: *M* = 3.26, *SD* = 1.37, α = 0.88; US sample: *M* = 3.24, *SD* = 1.50, α = 0.91). We measured Facebook political self-efficacy with six items adapted from Velasquez and LaRose (2015) such as "I am able to use Facebook to express my political views"; 1 = *strongly disagree*, 5 = *strongly agree*; German sample: *M* = 2.56, *SD* = 1.06, α = 0.92; US sample: *M* = 2.75, *SD* = 1.22, α = 0.95). We included these variables in the analyses to control for their influence.

In the survey, we did not present participants with a stimulus of a problematic user comment but asked for their general perception of who should intervene. Thus, we measured whether participants generally perceived user comments to be more of a benefit or a threat to society (1 = significant threat; 5 = significant benefit; German sample: *M* = 2.88, *SD* = 1.01; US sample: *M* = 2.93, *SD* = 0.96). We added this control variable to the analyses.

RESULTS

We computed four hierarchical linear regressions to test our hypotheses (Table 4). The participants' attribution of responsibility for intervention to the four actors served as dependent variables in the four regressions: (1) responsibility attribution to law enforcement,

TABLE 4 Hierarchical multiple regression analyses predicting attribution of responsibility to law enforcement, Facebook, news organizations, and self-responsibility from country of origin and support for free speech

	Responsibility attribution to law enforcement ^a (β)	Responsibility attribution to Facebook ^b (β)	Responsibility attribution to news organizations ^c (β)	Self-responsibility ^d (β)
Step 1				
Gender	-0.024	-0.005	0.066**	-0.018
Age	0.002	0.053*	0.091***	0.066**
Education ^e	-0.020	-0.050*	0.003	0.005
Political ideology ^f	-0.037	-0.101***	-0.090***	0.008
Step 2				
Facebook news use ^g	0.053*	0.030	0.040	0.125***
Facebook political self-efficacy ^h	0.100***	0.141***	0.140***	0.180***
Perception of user comments ⁱ	-0.030	-0.108***	-0.099***	0.004
Step 3				
Country of origin ^j	-0.278***	-0.068**	-0.091***	-0.060**
Support for free speech ^h	-0.061**	-0.010	0.042	0.002
<i>N</i>	2312	2312	2312	2312
Corr. R^2	0.090***	0.038***	0.045***	0.073***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

^a $F(9, 2302) = 26.441, p < 0.001$; Step 3: $\Delta R^2 = 0.069$, Sig. $\Delta R^2 p < 0.001$.

^b $F(9, 2302) = 11.132, p < 0.001$; Step 3: $\Delta R^2 = 0.004$, Sig. $\Delta R^2 p = 0.008$.

^c $F(9, 2302) = 17.962, p < 0.001$; Step 3: $\Delta R^2 = 0.010$, Sig. $\Delta R^2 p < 0.001$.

^d $F(9, 2302) = 21.364, p < 0.001$; Step 3: $\Delta R^2 = 0.003$, Sig. $\Delta R^2 p = 0.019$.

^e0 = Secondary education or less, 1 = secondary education or more.

^f0 = Liberal, 10 = conservative.

^g1 = Never, 6 = very often.

^h1 = Strongly disagree, 5 = strongly agree.

ⁱ1 = Significant threat, 5 = significant benefit.

^j0 = German, 1 = US American.

(2) to Facebook, (3) to news organizations, (4) self-responsibility. In Step 1 of the four regression models, we included the control variables age, gender, education, and political ideology. In Step 2, we included the participants' usage of Facebook for news, their self-efficacy to use Facebook for political purposes, and their perception of user comments. In Step 3, we introduced the independent variables country of origin and support for free speech. All variance inflation factors (VIF) were below the value of 1.7, indicating no significant multicollinearity issues.

The regression models show significant influences of the participants' country on attributing responsibility to the four actors. Germans attribute higher responsibility to all three institutionalized actors (0 = German, 1 = United States; Model 1: law enforcement: $\beta = -0.278, p < 0.001$; Model 2: Facebook: $\beta = -0.068, p = 0.002$; Model 3: news

organizations: $\beta = -0.091$, $p < 0.001$), supporting H1, H2, and H3. Germans also assume higher self-responsibility ($\beta = -0.060$, $p = 0.005$), answering RQ1.

Greater support for free speech has a significant negative influence on responsibility attribution to law enforcement ($\beta = -0.061$, $p = 0.004$), confirming H4a. It does not, however, influence responsibility attribution to Facebook ($\beta = -0.010$, $p = 0.654$, H4b), news organizations ($\beta = -0.042$, $p = 0.054$, RQ2), or self-responsibility ($\beta = 0.002$, $p = 0.938$, RQ3).

A *t* test showed that Germans display a significantly greater support for free speech ($n = 1155$, $M = 4.05$, $SD = 0.85$) than Americans ($n = 1164$, $M = 3.80$, $SD = 0.96$), $t(2288.338) = 6.583$, $p < 0.001$. This supports H5.

DISCUSSION

The goal of this study was to shed light on user attitudes toward a particular type of regulation of user-generated content: interventions against problematic comments on the social media pages of online news outlets. Through a survey, we uncovered attitudes about who users think should be in charge of interventions while also factoring in important antecedents of responsibility attribution: individual support for free speech as well as national background. We were interested in how Germans and Americans allocated responsibility for interventions against problematic comments to different actors: the state (law enforcement), platforms (Facebook), news organizations, and users themselves.

The data indicated medium-sized correlations between attributions of responsibility to the various actors: If people see a need for intervention, they more likely attribute the responsibility to do so to multiple actors. In both, the United States and Germany, in the ranking of responsibility attributions, Facebook came first, and news organizations came second. In Germany, this was followed by law enforcement, and then users, whereas in the US law enforcement came in last. This demonstrates the expectation for platforms and news outlets to take an overt role in managing problematic content.

When considering support for free speech, we found that this only negatively impacted responsibility attribution to law enforcement. This suggests that people who strongly support free speech are suspicious about state interference and fear potential harm to freedom of expression. However, support for free speech did not negatively affect responsibility attribution to Facebook. We can only speculate what may be underlying reasons: People may experience a strong need for regulation to an extent that it overwhelms possible skepticism against the platform operator. Perhaps people do not reflect on the power that said actor wields over free expression. It is up to future research to explore if this is a result of trust in Facebook or perceived helplessness toward online hate.

Moreover, support for free speech does not increase attribution of responsibility to the news organizations who maintain the Facebook pages. It seems that users do not perceive news organizations to have an explicit duty to control public discourse. Alternatively, it is worth investigating if users have internalized the legal norm that censorship only refers to restrictions by the state (not interventions by platform operators and news organizations). This could also explain why support for free speech was unrelated to attributions of responsibility to other actors but law enforcement (Chong, 1993).

Support for free speech does also not increase assumed self-responsibility of the users. This is in line with previous research that also did not find any impact of individual support for free speech on user interventions against uncivil comments (Naab, 2016; Naab et al., 2019). Those users who appreciate free speech probably object to intervention in general and emphasize a free exchange, even of harmful content. Alternatively, they might distrust the idea that users themselves are able or willing to maintain online discussions through their

self-regulatory power. A reason for this might be that, as user interventions address problematic content, their actions may also coincidentally leverage that very content they try to oppose. The reason is that (even disapproving) engagement with comments can increase their (algorithmic) visibility. This poses practical challenges to counter-speech in algorithmic settings.

When considering culture as a factor impacting responsibility attribution via the proxy of a participant being from Germany or the United States, we found that support for free speech differed substantially between the two countries. Germans displayed stronger support for free speech. This is congruent with previous survey studies (Wike et al., 2019; World Values Survey, 2019).

In line with prior research, we found that Germans attributed higher responsibility for interventions in comment sections to the state and its law enforcement agencies than Americans. This finding supports the notion that the German conception of free speech, which “merges liberal constitutionalism with a strong commitment to social solidarity” (Kommers, 2019, p. 561), is also reflected in the attitudes of the people. In tandem with stronger support for state intervention in Germany than in the United States (Nieuwenhuis, 2000), it is conceivable that Germans would favor interventions against online comments by law enforcement to a greater extent than Americans.

Our literature review led toward the assumption that Germans were more in favor of interventions by Facebook against uncivil comments on Facebook news pages. The empirical data shows that Germans attribute more responsibility to Facebook than Americans. It is possible that, due to a continuous debate about the Network Enforcement Act in Germany in the media and a “politicisation of the discussion” (Gollatz & Jenner, 2018, para. 1), Germans arrived at a position that assumes Facebook should take responsibility for what happens on its platform. The political debate about and media coverage of the law may have contributed to the public consciousness, so that it instigated deliberative processes among the German public about responsibility attribution in platform regulation.

In line with research finding higher media trust in Germany than in the United States (Newman et al., 2020), our study found that Germans attributed more responsibility for interventions to news organizations than Americans. This ties in with the conceptual idea that, at least in Germany, news outlets are perceived as responsible stewards of civic debate, in addition to their role as the Fourth Estate (Hampton, 2010). Future research needs to elaborate whether this difference is in fact tied to divergent levels of trust in news media, or if this relates to perceived power to impose sanctions and control public debates in social media.

We find that Germans assume greater self-responsibility in interventions against uncivil online comments than Americans. This is contradictory to what survey research suggests, namely that Americans are more convinced they have agential power to bring about change than Germans (Wike, 2016)—at least with regard to the self-regulatory power in online discussions. Perhaps, such faith to bring about change is contextual to self-actualization, and, therefore, less applicable to interventions against problematic speech online in the United States. In contrast, the results corroborate the observation that social movements like #iamhere, that promote user responsibility for counter-speech in online discussion, have a stronger basis in Germany than in the United States. Against the backdrop of public debates around the increasing challenges of handling uncivil comments, fake news, and social bots in social media, the task of intervening against problematic comments might seem overwhelming for platforms. Giving users more responsibility and agency to play their part may be a complementary and fruitful path forward.

LIMITATIONS

Our findings and interpretations must be seen in light of several limitations: We assumed various differences between Germans and Americans based on cultural heritage and aggregated survey data comparing these populations. We can only speculate if the differences that we found between Germans and Americans are indeed grounded in the arguments we make. Given the promising results, it seems a fruitful path for future research to measure the assumed influential factors on the individual level and use more comprehensive models to test the effects we show in this paper. This would entail measuring each participant's concept of free speech, a range of variables measuring national culture more holistically, trust in the state, platform operators, and news media as well as attitudes toward civic engagement. Furthermore, expanding the scope of future studies beyond Germany and the United States would yield interesting results beyond a western context of liberal democracy.

The survey did not provide respondents with a predetermined definition of the kinds of comments in need of intervention. The question left it up to the respondents' imagination which comments this might include. This allows the participants to make up their mind about comments they individually perceive as worthy of intervention. Yet it also allows for ambiguity in whether participants' ideas of what constitutes "offensive, inappropriate, or problematic" comments were comparable. Future studies could introduce stimulus-based designs that present preselected comments to participants. However, this will reduce the generalizability of the results and be further complicated by the need to translate stimuli between countries. Additionally, the study did not provide information as to what type of responsibility (beyond "intervention") the different regulatory actors should carry out. Future studies should provide more nuanced measurements for "attribution of responsibility." In combination with a differentiated measurement of the content of problematic comments, this will also allow investigating if participants expect varied types of interventions for different kinds of objectionable content. Finally, we focused on individual perceptions of online comments on Facebook because that is where news organizations increasingly locate their online discussions (Su et al., 2018). Expanding the reach to include and encompass other social media platforms would be a worthwhile endeavor for future research as well. These studies could also particularly consider perceptions of Internet users with lower education levels as these were slightly underrepresented in the current samples.

CONCLUSION

This study contributes to a better understanding of the preferences people have about who should be in charge of discussion spaces in two Western democracies. By reflecting on support for free speech in Germany and the United States and measuring perceptions of responsibility attribution in both countries, we contribute to a more comprehensive picture of user perceptions of regulatory actors involved in platform governance regimes. Specifically, we find that an individual's level of support for free speech impacts whether they see law enforcement in charge of comment regulation. While Germans support free speech more than Americans, they also think, more than Americans do, that the actors who should be responsible for taking care of problematic content are Facebook, the state, news organizations, and the users themselves. Against this background, nourishing freedom of expression in tandem with providing mechanisms of speech governance that allow users to intervene themselves are important next steps in thinking about the future of platform governance.

ACKNOWLEDGMENTS

The US portion of this project was funded through a grant from the William and Flora Hewlett Foundation awarded to the Center for Media Engagement in Moody College of Communication at The University of Texas at Austin, and the authors thank CME Director Natalie (Talia) Stroud for her support. The German portion of this project was funded by the Center for Media Convergence at the Johannes Gutenberg University of Mainz. Open Access funding enabled and organized by Projekt DEAL.

ENDNOTE

¹Dynata was formerly known as Research Now SSI and Survey Sampling International. Non-probability online panel surveys are no match in rigor to probability samples. Still, Dynata, as an established commercial operator, implements quality checks in their constitution of research populations for panels.

ORCID

Martin J. Riedl  <https://orcid.org/0000-0003-2411-1998>

Teresa K. Naab  <https://orcid.org/0000-0001-7345-2559>

Gina M. Masullo  <https://orcid.org/0000-0002-4909-2116>

Pablo Jost  <https://orcid.org/0000-0001-9267-1773>

Marc Ziegele  <https://orcid.org/0000-0002-2710-0955>

REFERENCES

- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The 'nasty effect': Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, 19(3), 373–387.
- Balkin, J. M. (2014). Old-school/new-school speech regulation. *Harvard Law Review*, 127, 2296–2342.
- Balkin, J. M. (2017). Free speech in the algorithmic society: Big data, private governance, and new school speech regulation. *UC Davis Law Review*, 51, 1149–1210.
- Brüggemann, M., Engesser, S., Büchel, F., Humprecht, E., & Castro, L. (2014). Hallin and Mancini revisited: Four empirical types of Western media systems. *Journal of Communication*, 64(6), 1037–1065.
- Caplan, R., & boyd, d. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data and Society*, 5(1), 1–12.
- Carmi, G. E. (2007). Dignity—The enemy from within: A theoretical and comparative analysis of human dignity as a free speech justification. *University of Pennsylvania Journal of Constitutional Law*, 9(4), 957–1002.
- Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. Palgrave Macmillan.
- Choi, Y. J., Choi, S., & Lee, M. J. (2012). Freedom of expression and the verification of identity policy in Korea. *Policy and Internet*, 4(3–4), 206–228.
- Chong, D. (1993). How people think, reason and feel about rights and liberties. *American Journal of Political Science*, 37(3), 867–899.
- Chong, D., & Levy, M. (2018). Competing norms of free expression and political tolerance. *Social Research: An International Quarterly*, 85(1), 197–227.
- Citron, D. K., & Wittes, B. (2017). The Internet will not break: Denying bad Samaritans Sec. 230 Immunity. *Fordham Law Review*, 86, 401–423.
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679.
- Corbin, C. M. (2009). The First Amendment right against compelled listening. *Boston University Law Review*, 89, 939–1016.
- Dahlberg, L. (2011). Re-constructing digital democracy: An outline of four 'positions'. *New Media & Society*, 13(6), 855–872.
- Daskal, E., Wentrup, R., & Shefet, D. (2020). Taming the Internet trolls with an Internet ombudsperson: Ethical social media. *Regulation Policy and Internet*, 12(2), 207–224.
- Duggan, M. (2017). *Online harassment 2017*. Pew Research Center. <https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/>
- Einwiller, S. A., & Kim, S. (2020). How online content providers moderate user-generated content to prevent harmful online communication: An analysis of policies and their implementation. *Policy & Internet*, 12, 184–206.
- Esau, K., Friess, D., & Eilders, C. (2017). Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet*, 9(3), 321–342.

- Fawzi, N. (2019). Untrustworthy news and the media as 'enemy of the people?' How a populist worldview shapes recipients' attitudes toward the media. *The International Journal of Press/Politics*, 24(2), 146–164.
- Golan, G. J., & Lim, J. S. (2016). Third-person effect of ISIS's recruitment propaganda: Online political self-efficacy and social media activism. *International Journal of Communication*, 10, 4681–4701.
- Gollatz, K., & Jenner, L. (2018). Hate speech and fake news—How two concepts got intertwined and politicised. *HIIG Digital Society Blog*. <https://www.hiig.de/hate-speech-fake-news-two-concepts-got-intertwined-politicised/>
- Gollatz, K., Riedl, M. J., & Pohlmann, J. (2018). Removals of online hate speech in numbers. *HIIG Science Blog*. <https://www.hiig.de/en/removals-of-online-hate-speech-numbers/>
- Hampton, M. (2010). The Fourth Estate ideal in journalism history. In S. Allen (Ed.), *Routledge companion to news and journalism* (pp. 3–12). Routledge.
- Hanitzsch, T., van Dalen, A., & Steindl, N. (2018). Caught in the nexus: A comparative and longitudinal analysis of public trust in the press. *International Journal of Press/Politics*, 23(1), 3–23.
- Hawdon, J., Oksanen, A., & Räsänen, P. (2017). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior*, 38(3), 254–266.
- Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, 34(1), 1–14.
- Heldt, A. (2019). Reading between the lines and the numbers: An analysis of the first NetzDG reports. *Internet Policy Review*, 8(2).
- Horwitz, R. B. (1991). The First Amendment meets some new technologies: Broadcasting, common carriers, and free speech in the 1990s. *Theory and Society*, 20(1), 21–72.
- Inglehart, R. (1997). *Modernization and postmodernization: Cultural, economic and political change in 43 societies*. Princeton University Press.
- Inhoffen, L. (2017). *Mehrheit der Deutschen findet Gesetzentwurf gegen Hasskommentare sinnvoll* [Majority of Germans find draft law against hateful comments sensible]. *YouGov*. <https://yougov.de/news/2017/04/15/mehrheit-der-deutschen-findet-gesetzentwurf-gegen-/>
- Janssen, D., & Kies, R. (2005, December). Online forums and deliberative democracy. *Acta Politica*, 40(3), 317–335.
- Kalch, A., & Naab, T. K. (2017). Replying, disliking, flagging: How users engage with uncivil and impolite comments on news sites. *Studies in Communication and Media*, 6(4), 395–419.
- Kang, C., & Isaac, M. (2020). U.S. and states say Facebook illegally crushed competition. *The New York Times*. <https://www.nytimes.com/2020/12/09/technology/facebook-antitrust-monopoly.html>
- Kenski, K., Coe, K., & Rains, S. A. (2017). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research*, 47(6), 795–814.
- Khan, R. A. (2013). Why do Europeans ban hate speech? A debate between Karl Loewenstein and Robert Post. *Hofstra Law Review*, 41(3), 545–585.
- Kim, Y. (2021). Understanding the bystander audience in online incivility encounters: Conceptual issues and future research questions. *Proceedings of the 54th Hawaii International Conference on System Sciences*. pp. 2934–2943.
- Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131, 1598–1670.
- Kohut, A., Wike, R., Horowitz, J. M., Poushter, J., Barker, C., Bell, J., & Gross, E. M. (2011). *Twenty years later: Confidence in democracy and capitalism wanes in former Soviet Union*. Pew Research Center. <https://www.pewresearch.org/wp-content/uploads/sites/2/2011/12/Pew-Global-Attitudes-Former-Soviet-Union-Report-FINAL-December-5-2011.pdf>
- Kommers, D. P. (2019). Can German constitutionalism serve as a model for the United States? *German Law Journal*, 20(4), 559–567.
- Kraut, R. E., & Resnick, P. (Eds.). (2012). *Building successful online communities: Evidence-based social design*. MIT Press.
- Krotoszynski, R. J. J. (2015). Free speech paternalism and free speech exceptionalism: Pervasive distrust of government and the contemporary First Amendment. *Ohio State Law Journal*, 76(3), 659–690.
- Ksiazek, T. B. (2015). Civil interactivity: How news organizations' commenting policies explain civility and hostility in user comments. *Journal of Broadcasting & Electronic Media*, 59(4), 556–573.
- Lambe, J. L. (2002). Dimensions of censorship: Reconceptualizing public willingness to censor. *Communication Law & Policy*, 7(2), 187–235.
- Landesanstalt für Medien NRW. (2020). "Ergebnisbericht: forsa-Befragung zu Hate Speech 2020." Düsseldorf: Landesanstalt für Medien NRW. https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Hass/forSa_LFMNRW_Hassrede2020_Ergebnisbericht.pdf
- Ley, H. (2018). *#ichbinhier: Zusammen gegen Fake News und Hass im Netz* [#iamhere: Together against fake news and hate in the net]. DuMont Buchverlag.

- Lim, J. S. (2017). The third-person effect of online advertising of cosmetic surgery: A path model for predicting restrictive versus corrective actions. *Journalism & Mass Communication Quarterly*, 94(4), 972–993.
- Lim, J. S., & Golan, G. J. (2011). Social media activism in response to the influence of political parody videos on YouTube. *Communication Research*, 38(5), 710–727.
- Livingstone, S. (2013). 'Knowledge enhancement': The risks and opportunities of evidence-based policy. In B. O'Neill, E. Staksrud, & S. McLaughlin (Eds.), *Towards a better internet for children: Policy pillars, players and paradoxes*. Nordicom.
- Lo, V. H., & Wei, R. (2005). Exposure to Internet pornography and Taiwanese adolescents' sexual attitudes and behavior. *Journal of Broadcasting & Electronic Media*, 49(2), 221–237.
- Maier, R. (2019). Self-responsibility: Transformations. *American Behavioral Scientist*, 63(1), 27–42.
- Masullo, G. M., Riedl, M. J., & Huang, Q. E. (2020). Engagement moderation: What journalists should say to improve online discussions. *Journalism Practice*. Advance online publication. <https://doi.org/10.1080/17512786.2020.1808858>
- Matias, J. N. (2019). The civic labor of volunteer moderators online. *Social Media and Society*, 5(2), 1–12.
- Medeiros, B. (2017). Platform (non-)intervention and the 'marketplace' paradigm for speech. *Regulation Social Media and Society*, 3(1), 1–10.
- Medeiros, B. (2019). The ideological significance of 'institutional neutrality' mandates in state-level campus speech legislation. *First Amendment Studies*, 53(1–2), 22–40.
- Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, 20(11), 4366–4383.
- Naab, T. K. (2016). Der Sanktionsbedarf von Facebook-Inhalten aus Sicht von NutzerInnen und seine Determinanten [The need to sanction Facebook content from the perspective of the users and its determinants]. *Medien & Kommunikationswissenschaft*, 64(1), 56–73.
- Naab, T. K., Heinbach, D., Ziegele, M., & Grasberger, M.-T. (2020). Comments and credibility: How critical user comments decrease perceived news article credibility. *Journalism Studies*, 21(6), 783–801.
- Naab, T. K., Kalch, A., & Meitz, T. (2018). Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on Bystander behavior. *New Media & Society*, 20(2), 777–795.
- Naab, T. K., Naab, T., & Brandmeier, J. (2019). Uncivil user comments increase users' intention to engage in corrective actions and their support for authoritative restrictive actions. *Journalism & Mass Communication Quarterly*. Advance online publication. <https://doi.org/10.1177/1077699019886586>
- Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2020). *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism, University of Oxford.
- Nieuwenhuis, A. (2000). Freedom of speech: USA vs. Germany and Europe. *Netherlands Quarterly of Human Rights*, 18(2), 195–214.
- Paek, H. J., Lambe, J. L., & McLeod, D. M. (2008). Antecedents to support for content restrictions. *Journalism & Mass Communication Quarterly*, 85(2), 273–290.
- Porten-Chée, P., Kunst, M., & Emmer, M. (2020). Online civic intervention: A new form of political participation under conditions of a disruptive online discourse. *International Journal of Communication*, 14, 514–534.
- Post, R. (2017). Legitimacy and hate speech. *Constitutional Commentary*, 32(3), 651–659.
- Puppis, M., & Van den Bulck, H. (2019). Introduction: Media policy and media policy research. In H. Van den Bulck, M. Puppis, K. Donders, & L. Van Audenhove (Eds.), *The Palgrave handbook of methods for media policy research* (pp. 3–21). Palgrave Macmillan.
- Riedl, M. J., Masullo, G. M., & Whipple, K. N. (2020). The downsides of digital labor: Exploring the toll incivility takes on online comment moderators. *Computers in Human Behavior*, 107(3), 106262.
- Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.
- Rojas, H., Shah, D. V., & Faber, R. J. (1996). For the good of others: Censorship and the third person effect. *International Journal of Public Opinion Research*, 8(2), 163–186.
- Schulz, W. (2018). Regulating intermediaries to protect privacy online: The case of the German NetzDG. *HIIG Discussion Paper Series*. <https://papers.ssrn.com/abstract=3216572>
- Shen, F. (2017). Internet use, freedom supply, and demand for Internet freedom: A cross-national study of 20 countries. *International Journal of Communication*, 11, 2093–2114.
- Stroud, N. J., Scacco, J. M., Muddiman, A., & Curry, A. L. (2015). Changing deliberative norms on news organizations' Facebook sites. *Journal of Computer-Mediated Communication*, 20, 188–203.
- Su, L. Y. F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society*, 20(10), 3678–3699.
- Thompson, D. F. (2014). Responsibility for failures of government: The problem of many hands. *The American Review of Public Administration*, 44(3), 259–273.
- Velasquez, A., & LaRose, R. (2015). Youth collective activism through social media: The role of collective efficacy. *New Media & Society*, 17(6), 899–918.

- Wei, R., & Lo, V. H. (2007). The third-person effects of political attack ads in the 2004 U.S. presidential election. *Media Psychology*, 9(2), 367–388.
- Wike, R. (2016). *5 ways Americans and Europeans are different*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2016/04/19/5-ways-americans-and-europeans-are-different/>
- Wike, R., Poushter, J., Silver, L., & Cornibert, S. (2019). *European public opinion three decades after the fall of communism*. Pew Research Center. <https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/10/Pew-Research-Center-Value-of-Europe-report-FINAL-UPDATED.pdf>
- Wike, R., Simmons, K., Stokes, B., & Fetterolf, J. (2017). *Globally, broad support for representative and direct democracy*. Pew Research Center. https://www.pewresearch.org/global/wp-content/uploads/sites/2/2017/10/Pew-Research-Center_Democracy-Report_2017.10.16.pdf
- Wise, K., Hamman, B., & Thorson, K. (2006). Moderation, response rate, and message interactivity: Features of online communities and their effects on intent to participate. *Journal of Computer-Mediated Communication*, 12(1), 24–41.
- World Values Survey. (2019). World Value Survey Wave 6: 2010–2014. <https://www.worldvaluessurvey.org/WVSOnline.jsp>
- Wright, S. (2006). Government-run online discussion fora: Moderation, censorship and the shadow of control. *The British Journal of Politics and International Relations*, 8(4), 550–568.
- Wright, S., Jackson, D., & Graham, T. (2019). When journalists go 'below the line': Comment spaces at The Guardian (2006–2017). *Journalism Studies*, 29(4), 1–20.
- Ziegele, M., & Jost, P. (2020). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research*, 47(6), 891–920.
- Ziegele, M., Koehler, C., & Weber, M. (2018). Socially destructive? Effects of negative and hateful user comments on readers' donation behavior toward refugees and homeless persons. *Journal of Broadcasting & Electronic Media*, 62(4), 636–653.
- Ziegele, M., Naab, T. K., & Jost, P. (2019). Lonely together? Identifying the determinants of collective corrective action against uncivil comments. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/1461444819870130>

How to cite this article: Riedl, M. J., Naab, T. K., Masullo, G. M., Jost, P., & Ziegele, M. (2021). Who is responsible for interventions against problematic comments? Comparing user attitudes in Germany and the United States. *Policy & Internet*, 1–19. <https://doi.org/10.1002/poi3.257>