Article

"Intervening Is a Good Thing but . . .": The Role of Social Norms in Users' Justifications of (Non-)Intervention Against Incivility

Emilija Gagrčin[1,2] 🆔 and Miriam Milzner[2]

Abstract

User intervention against incivility as social enforcement of democratic norms on social media platforms is considered an act of "good citizenship" by citizens and scholars alike. However, between ideals and behavior, multiple social norms are at play in shaping individuals' sense of personal responsibility for intervening. This study explores the role of conflicting norms in situations requiring user intervention against online incivility. By combining the perspectives of norms as expectations and norms as cultural vocabularies, we investigate users' salient norms, and how these norms influence users' justifications for (non-)intervention. Based on qualitative interview data from Germany ($N = 20$), we identified three distinct reasoning patterns employed to justify (non-)intervention: *the pragmatic, the dismissive, and the aspirational*. By identifying fault lines, our typology points to normative origins of ambivalence related to user intervention. The findings offer insights into strategies to motivate intervention against online incivility.

Keywords

incivility, user intervention, counterspeech, good citizenship, social norms

Online discussions in which groups or individuals are belittled, threatened, or insulted can be understood as digital emergencies that threaten individuals, groups, and democratic values (Naab, 2016). In theory, when witnessing such an emergency, users are faced with deciding whether to intervene—a decision that is influenced by factors such as their perception of the urgency of the situation and their sense of responsibility to act (Latané & Darley, 1970; Leonhard et al., 2018; Naab, 2016). Initiatives[1] that foster a sense of responsibility typically communicate core civic duties and virtues such as defending democracy, practicing solidarity, and showing civil courage, and assert that countering hate speech online requires civil courage. Seeking to influence the normative frameworks that people draw on to understand their roles in specific situations (Swidler, 1986; Thorson, 2012), they typically highlight the significance of incivility as a violation of norms and emphasize the desirability and democratic value of user intervention (Cialdini et al., 1991; Stok & de Ridder, 2019). On a positive note, many individuals recognize that intervention against incivility online is a manifestation of good citizenship (Emmer et al., 2021; Heger et al., 2022). However, the actual occurrence of intervention in response to such incidents remains limited to a minority

(Emmer et al., 2021; Heger et al., 2022), prompting questions about the reasons for this discrepancy and the justifications individuals provide for possible incongruencies in their behavior.

The present study explores the role of conflicting norms in situations requiring user intervention. In an instance of online incivility in the context of everyday social media use, users may feel torn between the injunctive norm of promoting civility (Kunst et al., 2021; Ziegele et al., 2020) and the risk of escalating conflict through personal attacks, which users may believe is typical (Gagrčin, 2022; Shmargad et al., 2022). In other words, such situations involve multiple and potentially competing norms (Chung & Rimal, 2016; Cialdini et al., 1991; Kallgren et al., 2000). Understanding users' justifications for intervening or refraining from doing so is vital in creating effective strategies to encourage

[1]University of Mannheim, Germany
[2]Weizenbaum-Institute for the Networked Society, Germany

Corresponding Author:
Emilija Gagrčin, Institute for Media and Communication Studies, University of Mannheim, B 6, 30–32, 68131 Mannheim, Germany.
Email: gagrcin@uni-mannheim.de

intervention against online incivility. Thus, we combine the perspectives of norms as expectations (Cialdini et al., 1991; Rimal & Real, 2003) and norms as part of cultural vocabularies (Swidler, 1986; Thorson, 2012) to empirically investigate (a) which norms social media users perceive and focus on concerning intervention against incivility and (b) how these norms matter in users' justifications of (non-)intervention. The empirical analysis is based on the data set from vignette interviews with 20 social media users (aged 18–25 years) in Germany.

Literature Review

Injunctive and Descriptive Citizenship Norms

Social norms are informal rules that regulate social life by eliciting conformity (Chung & Rimal, 2016, p. 4). The present study is interested in social norms concerning everyday political talk on social media platforms. These are typically studied under the umbrella of citizenship norms and defined as expectations toward civic and political behavior (Dalton, 2008; Kligler-Vilenchik, 2017). As social norms, citizenship norms are learned through various forms of mediated and non-mediated communication (Chung & Rimal, 2016; Gagrčin et al., 2022). They differ from *good citizenship ideals* that are typically defined by scholars as a coherent perspective from which citizens' behavior can be evaluated as desirable, necessary, and legitimate, based on a higher common good such as democracy or a public sphere (Hove, 2021, p. 894). As such, good citizenship ideals lack specific guidance for concrete situations and are, at best, theoretical and abstract expectations (Bormann et al., 2022; Lindenberg, 2008). Empirical research shows that this type of normative insight alone is often insufficient to motivate action (Gagrčin, 2022; Thorson, 2015). Instead, the Focus Theory of Normative Conduct suggests that people rely on injunctive and descriptive norms to inform their sense of responsibility and intention to act in specific situations (Cialdini et al., 1990; Jacobson et al., 2011; Stok & de Ridder, 2019).

In line with the literature on normative influence (Chung & Rimal, 2016; Legros & Cislaghi, 2020), we can differentiate between injunctive and descriptive norms related to political behavior on social media. Injunctive norms, also known as the norms of *ought* (Cialdini et al., 1991), stand for the pressure individuals feel to engage in certain behaviors based on the expectations of others (Chung & Rimal, 2016; Rimal & Real, 2003). Theoretically, injunctive norms can also entail disapproval of deviant behavior (Chung & Rimal, 2016). Thus, injunctive norms manifest an *inter*personal character, as they serve to achieve social approval and avoid social sanctions (Jacobson et al., 2011). They may also have a collective character, conveying expectations toward behaviors considered desirable for collective goods such as public discourse (Gagrčin & Porten-Cheé, 2023). In such cases, they can be seen as a translation of ideals into social expectations on the

individual level. On the contrary, descriptive norms, or the norms of *is* (Cialdini et al., 1991), are beliefs about how typical or expectable a behavior is in a given situation, thus conveying information about appropriate or correct ways to behave (Chung & Rimal, 2016; Rimal & Real, 2003). Accordingly, they are said to have an *intra*personal character and provide individuals with cognitive shortcuts to determine the most effective and efficient course of action (Jacobson et al., 2011; Stok & de Ridder, 2019). In the following, we briefly assess incivility and user intervention according to the norms approach.

Incivility and User Intervention: Norm Violation and Enforcement

Incivility in online discussions refers to user comments that belittle, threaten, or insult individuals and groups (Naab, 2016; Naab et al., 2021), which violates deliberative ideals and injunctive norms of civility (Bormann et al., 2022; Rossini, 2022). As such, uncivil comments are likely to attract sanctions such as social disapproval or counterspeech (Porten-Cheé et al., 2020). Although platforms offer possibilities to formally sanction uncivil comments by reporting them, they are also known for their problems with removing reported content (Gillespie, 2020). Thus, counterspeech by social media users is argued to be a powerful corrective in the online discourse (Friess et al., 2021; Leonhard et al., 2018; Porten-Cheé et al., 2020). Here, users who engage in intervention against incivility act as norm enforcers by encouraging conformity and maintaining citizenship norms (Legros & Cislaghi, 2020).

Since user intervention against incivility is more likely when people perceive a situation as urgent and feel personally responsible for intervention (Latané & Darley, 1970), research in this area has typically investigated individual perceptions of uncivil content (Leonhard et al., 2018; Naab et al., 2018, 2021) and users' sense of responsibility stemming from their observations of and experiences with incivility online (Gagrčin et al., 2022; Kunst et al., 2021; Ziegele et al., 2020). However, less attention has been paid to perceptions of intervention as norm enforcement that is itself subject to norms (Fehr & Fischbacher, 2004), and how these norms may be at odds with one another, shaping ideas about responsibility. Thus, in the following, we consider descriptive and injunctive norms related to incivility and intervention and their relevance in user intervention.

Norms in User Intervention

Although descriptive and injunctive norms frequently overlap, they can also conflict and have different consequences (Chung & Rimal, 2016) for the perception of urgency and the responsibility for intervention. Specifically, when drawing on injunctive norms, people are less concerned with finding

the most effective behavioral decision but rather with achieving a more distant goal (Stok & de Ridder, 2019). Unlike injunctive norms, descriptive norms are more directly related to the behavior at hand since people use the behavior of others as a cue for adaptive behavior (Stok & de Ridder, 2019). For example, in some online spaces, uncivil behavior can be considered typical—albeit undesirable—and, in turn, model uncivil behavior (Álvarez-Benjumea & Winter, 2018). At the very least, descriptive norms of incivility may lessen a sense of urgency and personal responsibility for intervening. At the same time, while some users are also more inclined to intervene in online spaces where intervening behavior is widespread (i.e., descriptive norm) (Buerger, 2021; Friess et al., 2021; Miškolci et al., 2020), research shows that intervention becomes *less* likely in the presence of a high number of bystanders (Leonhard et al., 2018). The latter suggests that some people may feel less personally responsible if they can rely on others to do the job. In line with this, repeated surveys show that although most respondents report encountering incivility online and believing one should do something about it, only a minority effectively intervenes against it (Emmer et al., 2021; Heger et al., 2022).

The Focus Theory of Normative Conduct states that the most influential norm for actual behavior in situations with multiple norms is the focal norm, which is the one made salient and on which attention is focused (Jacobson et al., 2011; Stok & de Ridder, 2019). Situational factors, including personal involvement, may determine which norm becomes focal (Lapinski & Rimal, 2005). For example, people feel a stronger expectation to intervene against problematic online behavior from close social ties compared with strangers, likely due to negative stereotypes and the perception that engaging with uncivil or deviant strangers is not worthwhile (Gagrčin, 2022; Neubaum et al., 2021). In addition, the cognitive resources available to individuals in a given situation influence which norms become salient (Jacobson et al., 2011). For instance, when individuals experience limited cognitive capacity due to tiredness or multiple attention-demanding tasks, which is common in social media use (Mark, 2023), they tend to rely on cognitive shortcuts and prioritize descriptive norms (Jacobson et al., 2011). Conversely, individuals with higher cognitive capacity or those motivated to reflect on their values and self-concepts are more likely to consider distant goals and act according to injunctive norms (Jacobson et al., 2011; Mark, 2023; Stok & de Ridder, 2019). Since injunctive and descriptive norms are associated with different goals, they direct individuals' focus on different aspects of themselves, influencing how individuals perceive their responsibility to act and their behavioral intentions (Jacobson et al., 2011; Kallgren et al., 2000).

Specifically, our study aims to explore how perceptions of different norms matter for individuals' justification of their decision to intervene in situations of incivility on social media. To achieve this, we adopt the concept of normative vocabularies, conceived as a pool of resources that people use to interpret situations and formulate appropriate responses (Swidler, 1986; Thorson, 2012). In this sense, normative vocabularies contain both perceived injunctive and descriptive expectations toward certain behaviors, traits, and situations as well as more abstract ideals. As such, the concept of vocabularies is beneficial in studying users' justifications due to its flexibility: It enables us to understand how people use normative and attentional resources available to them to make sense of their participation in public life, how they develop a personal philosophy of their role, and how this informs their actions and expectations of others (Gagrčin & Porten-Cheé, 2023; Thorson, 2012). Thus, we seek to answer two research questions:

> What citizenship norms do social media users perceive regarding a) incivility and b) intervention against incivility in comment sections on social media (RQ1)? How are these norms used to justify responsibility for intervention (RQ2)?

Method

Study Context

This study is situated within the German Facebook context, which has garnered significant scholarly attention regarding user intervention (Friess et al., 2021; Kalch & Naab, 2017; Kunst et al., 2021; Naab, 2016; Naab et al., 2018; Ziegele et al., 2020). The interest in user intervention among Germans can be attributed to the societal significance placed on civil courage, where individuals assist others at personal risk (Willems, 2021, p. 679). User intervention against incivility aligns with this notion of civil courage and is reinforced by various educational and civil initiatives like the online movement #ichbinhier, klicksafe.de, and hass-im-netz.info. The societal value placed on combating incivility and hate speech in Germany is exemplified by the "Network Enforcement Act" enacted by the German Parliament in 2017. This legislation mandates social media platforms to promptly remove content that violates existing laws on unlawful communication or face substantial fines (Tworek & Leerssen, 2019). The Network Enforcement Act covers, for example, defamation, hate speech, and Holocaust denial. It relies heavily on user awareness and engagement in reporting instances of hate speech and incivility, underscoring the importance of citizens' norm perceptions.

We study Facebook because, in a recent study, users reported Facebook as the most likely space where they encounter incivility (Reichelmann et al., 2021). Moreover, Facebook is one of Germany's most-used social media platforms (Newman et al., 2022). To narrow the scope of inquiry, our main interest lies in user engagement in news outlets' comments on social media posts as one of the prominent arenas in the online public sphere that is in particular need of user intervention (Leonhard et al., 2018; Ziegele et al., 2020).
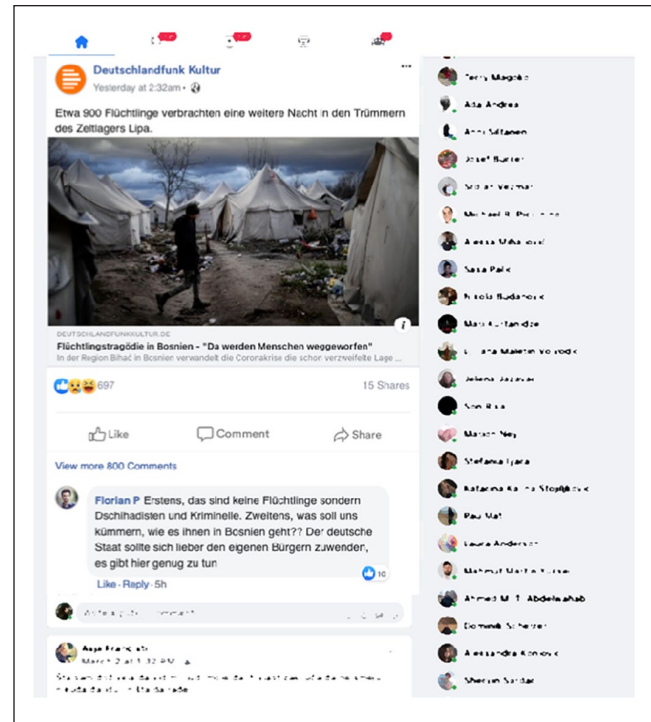
Table 1. Participants (Pseudonymized).

Name	Gender	Age	Area of studies
Anouk	F	20	Humanities
Inga	F	21	Humanities
Bjarne	M	18	Natural sciences
Claudia	F	25	Natural sciences
Karlo	M	24	Natural sciences
Luis	M	21	Natural sciences
Maik	M	22	Natural sciences
Niklas	M	18	Natural sciences
Theresa	F	25	Natural sciences
Alina	F	25	Social sciences
Astrid	F	21	Social sciences
Constantin	M	25	Social sciences
Else	F	25	Social sciences
Elvira	F	22	Social sciences
Jasmin	F	25	Social sciences
Lennard	M	20	Social sciences
Lidia	F	24	Social sciences
Maren	M	22	Social sciences
Till	M	25	Social sciences
Zoe	F	24	Social sciences

Figure 1. Vignette.
Note. Article title (translated): "Refugee tragedy in Bosnia: 'People are being thrown away'"; post caption: "About 900 refugees spent another night in the ruins of the Lipa tent camp"; Florian P's post: "First, these are not refugees but jihadists and criminals. Second, why should we care how they are doing in Bosnia? The German state should better turn to its own citizens. There is enough to do here." The participants read the following text attached to the picture: "While scrolling through his Facebook, Stefan (21, student) stumbled upon a post by Deutschlandfunk Kultur about the situation in the refugee camp Lipa in Bosnia and Herzegovina. In the comment section, he saw a comment by Florian P."

Participants

We employed convenience sampling and recruited participants via mailing lists of the major universities in town. To ensure some diversity in the sample, potential participants filled in a prescreening survey answering basic demographic questions (age, gender, field of study) and a few questions specific to the interest of the study (political interest, use of social media, frequency of encounters with incivility and frequency of intervention). Participants were offered a gift voucher worth 15€ as an incentive for participation. The final sample counts 20 participants aged between 18 and 25 years (23 was the average age), with a slightly lower representation of natural sciences students (35%) and participants who self-identified as male (45%) (see Table 1). However, the sample was diverse regarding intervention frequency, and we aimed to avoid overrepresentation by intervention enthusiasts. All participants had at least a passive Facebook profile and were daily users of multiple social media platforms.

Interviews

We employed vignette interviews as a standalone method to explore individual normative frameworks in situations requiring user intervention. Vignettes are hypothetical or fictional stories presented to participants, accompanied by questions (Gray et al., 2017). This approach helps mitigate social desirability biases commonly encountered when eliciting norms in interviews. The vignette used in this study portrayed a post from a prominent German radio station on their Facebook Page, reconstructed using the online tool Zeoob (Figure 1). The post included a linked article about poor conditions in a refugee camp in Bosnia, accompanied by a brief caption and an uncivil yet polite user comment. Impoliteness involves breaches of etiquette, such as name-calling and offensive language, and is easier to detect than incivility cloaked in politeness (Kalch & Naab, 2017; Papacharissi, 2004). Since we are interested in the subversion of democratic norms of civility and users' engagement in upholding those norms, it was essential to ensure that participants focus on the discriminating dimension of the post.

We asked participants to consider situations both from ideal and pragmatic points of view, i.e., thinking about what is ultimately desirable (norms as ideals), what they feel is socially expected (injunctive norms), and what is likely to happen based on their beliefs about typical behaviors in such situations (descriptive norms) (Finch, 1987; Gray et al., 2017). All interviews took place virtually and lasted about 1 hour. Only the sound was recorded.
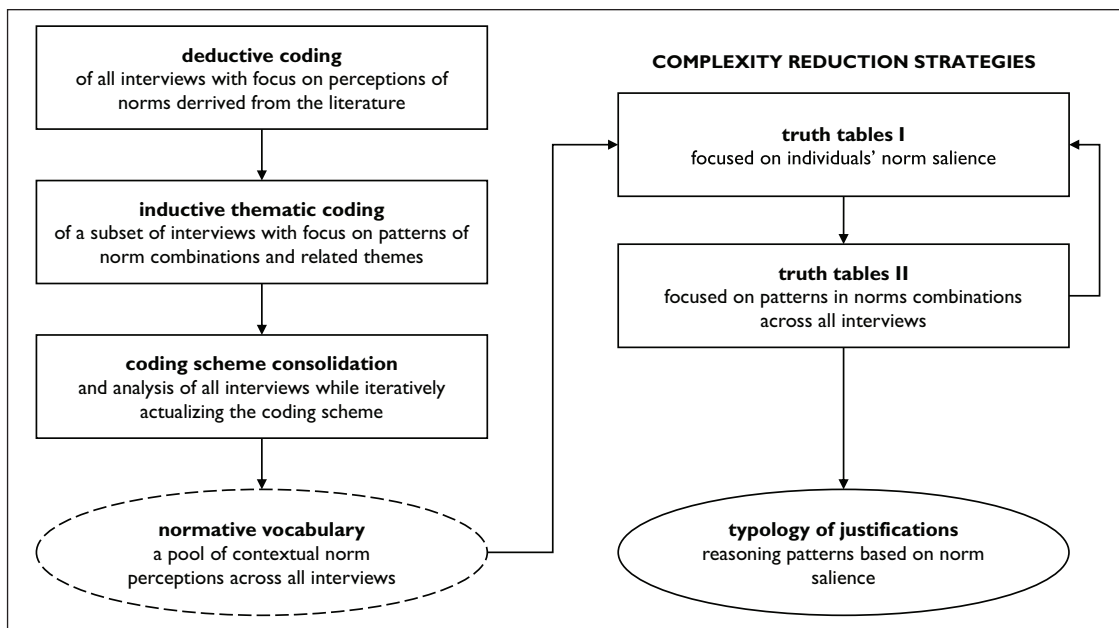
Figure 2. Analytical steps.

Table 2. Normative Structure of Justifications.

	Vocabularies		
	Pragmatic	Aspirational	Dismissive
Focal norms			
Descriptive norms	+++	+	−
Injunctive norms	+	+++	−−
Ideals	++	++	−−−
Consequences	Abstract expectations toward intervention	Abstract and specific expectations toward intervention	No expectations toward intervention

Note. The +/− used illustrate the weighting that the respective vocabularies assign to norms and ideals. The number indicates the extent to which this norm is referred to justify (non)intervention. The + indicates the extent of affirmative reference, and the − describes the extent of rejection of this norm.

Analysis

Both authors analyzed the generated data. Initially, we employed a deductive approach to investigate participants' perceptions of norms related to incivility and user intervention (RQ1). We developed a comprehensive list of codes encompassing descriptive and injunctive norms and norms as ideals. This deductive coding process was applied to all interviews. In the next phase, a subset of interviews was exploratively analyzed (each author examined five interviews) to understand the relevance of identified norms for participants' justifications of (non-)intervention (RQ2). These initial findings were a foundation for refining the coding scheme and analyzing the remaining interviews. The themes and codes were iteratively adjusted based on the data and theoretical insights. This process resulted in a collection of perceived norms and associated themes, reflecting shared contextual perceptions of social norms and themes such as fatigue and conflict avoidance.

While consolidating the coding scheme and establishing the normative vocabulary, we noticed a significant divergence in how the vocabulary and perceived norms were individually treated and interpreted. This realization prompted us to explore and uncover previously unnoticed relationships and connections. As a result, we revisited the data using truth tables to reduce complexity (Ragin & Amoroso, 2011). Initially, our focus was on individuals' perception of norm salience. However, we soon recognized the insufficiency of this emphasis on individuals. In subsequent iterations, we shifted our attention to distinct patterns of norm combinations that serve to justify (non)intervention. This involved integrating different norm perceptions, participants' focal norms, and the resulting implications for responsibility. The analytical procedure is illustrated in Figure 2. By considering these factors together, we derived a typology of intervention justifications (Table 2).

Findings

Perceptions of Incivility and Uncivil Users

Participants were not provided with a specific definition of incivility. Instead, at the start of the interview, they were asked to describe instances of incivility they had observed recently. Despite the lack of a formal definition, participants consistently characterized incivility as content that "devalues and dehumanizes people" (Leonie) and posts that are not "just an opinion, but an insult" (Rebeca). As demonstrated in other studies (Emmer et al., 2021; Savimäki et al., 2020), seeing incivility online was very common in the sample. Participants were more likely to see incivility toward people exposed on social media platforms, such as people in politics. Examples cited by participants included religious devaluation, xenophobia, and incitement against LGBTQ+ individuals and women. However, participants noted that incivility appeared less prevalent in social media spaces centered around art, culture, and sociability. Overall, the normalization of uncivil behavior on social media was evident, as expressed by Claudia's statement: "Incivility is just ubiquitous. There are so many comments like that. I have kind of gotten used to it."

None of the participants indicated having posted a hateful or uncivil comment online; when prompted, they firmly rejected the notion that incivility could ever be justified. Thus, throughout the interviews, participants consistently perceived incivility as a norm violation and expressed disapproval. For example, it was common to hear that "those [dehumanizing] positions are despicable" (Astrid) and descriptions of incivility as "very, very bad and terrible" (Maren). In echoing previous research (Buerger, 2021; Ziegele et al., 2020), participants also believed such behavior was harmful to the public discourse. Rebeca captures this sentiment by stating: "There are many people with whom I disagree, but in real life, you must listen to each other in some way. Now that everything has moved online, I'm worried about what that does to a democracy."

Notwithstanding their disapproval of uncivil *behavior*, many participants were sympathetic to perpetrators of incivility. Stereotypically, participants imagined uncivil users as individuals who are frustrated or "stuck" because of a personal history of living in precarity or growing up surrounded by a "wrong ideology." For example,

> There are enough times when I think, "What an intolerant asshole." But at the same time, I don't know that person's story, which maybe explains what they say. It doesn't *justify* what they do but it may explain it. (Else)

> I condemn their way of thinking. But it's probably not completely their fault; they didn't come up with it on their own. (Niklas)

Perception of perpetrators as disadvantaged individuals allowed participants to empathize with uncivil behavior while maintaining the overarching narrative that incivility is to be condemned and one should always behave civilly.

Perceptions of Intervention and Intervening Users

Contrary to incivility, estimating the prevalence of user intervention lacked coherence. Some participants thought that "someone always does it," while others insisted that most users only watch and "no one intervenes." That participants' estimates differed is less interesting since both are possible: while most people, indeed, do not intervene (e.g., Emmer et al., 2021), it may be that *enough* people intervene to create an impression that intervention is common (Friess et al., 2021). More importantly, the difference in estimates suggests that participants had a different cognitive focus, in the sense that some emphasized incivility (norm violation) while others focused on intervention (norm enforcement). This discrepancy in focus also influenced participants' perceptions of those who intervened. Generally, participants believed that a specific type of person was more likely to intervene. Some viewed these users as self-assured and respectable, while others regarded them "just as annoying as those who post uncivil comments" (Alina), or as individuals with excessive free time, according to Till.

On the injunctive level, the intervention was typically—though not exclusively—characterized as aspirational and admirable. Based on the perceptions of harm that incivility induced on the societal level, it was widely held that one ought to do something because "the perpetrator should see that he belongs to a minority and that his opinion is not right . . . that his opinion is anything but right" (Bjarne). Although common, statements like this one mostly remained at the level of ideals as abstract expectations:

> When it comes to how we live together, it just makes sense that we shouldn't tolerate things like incivility. In real life, when someone is attacked, we all feel like we should step in and do something. So, I think that same idea should apply online too. (Astrid)

At the same time, and based on their experiences and observations, participants held that interventions typically end in personal attacks, fights, and unconstructive exchanges. All of this made them highly aware that intervention of any kind requires resources such as time, emotional capacity, or self-confidence:

> You need a lot of self-confidence . . . you must be able to handle it well psychologically. Because intervention is not all that easy. Even if it's online, it is still a burden. (Lennard)

> I intervened a couple of times, and it escalated every time. At some point, I thought, "I don't want to put up with this anymore." (Else)

Note that participants focus their attention inwardly, using their anticipated emotional states motivated by descriptive

expectations as shortcuts to determine the most effective and efficient course of action for them (Jacobson et al., 2011; Stok & de Ridder, 2019). In turn, and since most participants believed that "intervention is a good thing to do," the space between good citizenship ideals and injunctive expectations was frequently populated with good reasons not to intervene or condemn non-intervention.

A Typology of Users' Justifications Related to Intervention Against Incivility

Despite a shared pool of perceived norms, participants drew different conclusions about their responsibilities and intentions to intervene. Drawing from the works of Swidler (1986) and Thorson (2012), who emphasize the significance of how people combine norms and resources to formulate "good enough" reasons for action, we identified three justifications for (non-)intervention: *pragmatic, dismissive*, and *aspirational*. Rather than characterizing types of people, these three justifications refer to the reasoning patterns that focus on different expectations to interpret situations of incivility, draw conclusions about user accountability, and formulate appropriate responses (Table 2).

Even if, theoretically, these justifications could be used by the same person in different moments or when faced with different types of incivility, we found that participants tended to use a corresponding vocabulary fairly consistently in their reasoning process. The pragmatic vocabulary was used most frequently, while fewer individuals used aspirational and dismissive justifications.

Aspirational: "Standing by Makes You Complicit." If scholars were to look for a "good intervening citizen," they would likely be very content with participants who employed the aspirational vocabulary. Aspirational vocabulary focuses on incivility as a norm violation that makes online intervention necessary and worthwhile. Injunctive norms, manifesting in individuals' willingness to condemn a behavior because it does not live up to the expectations of socially desirable behavior, are focal norms in this vocabulary. Consequently, the aspirational vocabulary values intervention and condemns non-intervention.

In line with previous studies (Gagrčin & Porten-Cheé, 2023; Kunst et al., 2021; Ziegele et al., 2020), this vocabulary foregrounds solidarity and altruism, which enables people to transcend immediate personal matters when thinking about intervention and instead focus on more distant goals (Fowler & Kam, 2003; Robison, 2022). Accordingly, the aspirational vocabulary disapproves of non-intervention in general and people who do not intervene in particular. As participants explain,

> It's just no effort to write something and to show people that it's not true, that it's just bs what [uncivil users] say. It's just a matter of two minutes, and that's why I don't think it's okay to tolerate [incivility]. (Luis)

On the one hand, people who don't intervene are at least not the people who spread such things, but on the other hand, they become complicit because they don't position themselves against it openly. (Franziska)

Thus, aspirational vocabulary exhibits abstract and specific expectations toward intervention, and descriptive incivility norms reinforce the sense of responsibility. Grounded in what Hove (2021, p. 888) terms deontological reasoning, participants who employed this vocabulary had an identified "higher common good" in mind that transcended their personal interests, namely a non-discriminatory public discourse (Buerger, 2021; Gagrčin et al., 2022; Ziegele et al., 2020).

Notwithstanding the aspirational tone, these participants did not claim to intervene always. They acknowledged the pitfalls of intervention, such as high emotional costs. Nevertheless, the internally consistent perspective enabled them to bridge their position and capacities with the public interest, allowing them to make coherent judgments and criticize others' deficient justifications (Hove, 2021):

> I get it if you don't want to do something because you're worried about your safety. But if you're just not doing anything because it's easier and you're thinking, "Oh, someone else will handle it," then I think that's petty. (Astrid)

Dismissive: "Social Media Discourses Are Irrelevant." The dismissive vocabulary reflects a complete rejection of the importance of citizenship norms in the digital realm without a specific focus on any particular norm. This vocabulary was commonly used by participants who viewed the internet and social media as irrelevant spaces, resulting in their dismissal of both abstract and specific expectations regarding user intervention. This finding aligns with existing research indicating that individuals with negative perceptions of social media as a political platform are less inclined to engage in political participation through these technologies (Kwak et al., 2018). The dismissive vocabulary prominently exhibits a rejection or downplaying of the responsibility to intervene:

> It won't make any difference whether one intervenes by writing a comment. It's not an effective measure to do something good for society. So, there's no imperative to intervene from my point of view. (Maik)

This vocabulary features negative—at times even cynical—remarks related to intervening users, implying egoistic and ingenuine motives behind a presumably noble behavior:

> I think intervention has to do with selfish reasons. People just want to feel better about themselves. I guess if that's really important to you and you really want to do something for the democratic order, then you don't do it on social media. Social media has a lot to do with your ego and wanting to feel good. (Inga)

However, this does not mean that the dismissive vocabulary denies the importance of intervention altogether. Instead, it emphasizes other meaningful ways to counteract incivility and hate in society: "I'd rather go to a protest or get involved in some other more meaningful way offline, but not on the internet" (Maik). By redirecting the relevance of intervention to the offline realm, dismissive justifications displace personal responsibility for intervention in social media environments.

Pragmatic: "Intervening Is a Good Thing, but . . ..." While recognizing the problematic nature of incivility and the general responsibility for countering it, the pragmatic vocabulary primarily focuses on descriptive norms related to intervention and the typical unfolding of intervention situations. It also centers on the personal consequences and costs associated with the intervention. Participants who used the pragmatic vocabulary often shared how their beliefs about user intervention had changed over time. They may have attempted intervention in the past, but they now recognize the need to be more cautious with their resources. Alina explained, "I've noticed in the past how uncivil comments would upset and demoralize me. I realized that reading such content triggers negative emotions, so I consciously chose to disengage from it." Although this vocabulary implies a general responsibility to intervene, it does not translate into specific expectations for intervention, unlike the aspirational vocabulary. Instead, the descriptive norms around intervention (such as that it ends in conflict and may take a toll on one's mood) shift the focus to intrapersonal goals and allow for the acceptance of non-intervention.

Furthermore, pragmatic justifications implied that though noble in intentions, interventions against incivility were, in fact, neither meaningful nor efficient and required emotional and time investment. More than the other two vocabularies, the pragmatic vocabulary was marked by a sense of fatigue and habituation in the face of uncivil and deviant behavior on social media:

> It feels like there's rarely any meaningful dialogue about any topic. Everyone already has their fixed opinions and just resorts to insulting each other. And when most of the comments are negative, and sometimes I actually expect them to be, I get tired and don't bother reporting or even looking at them. (Anouk)

> Intervention is hardly ever welcomed with open arms. Even if you meant well and wanted to bring attention to something or simply express a different opinion, you might end up suffering for it. (Theresa)

Many participants who employed this vocabulary feared personal attacks, often emphasizing their conflict aversion. Bjarne told us he knows "many people who don't have social media because they don't want the stress of being misunderstood." This is consistent with previous research (Savimäki et al., 2020; Vraga et al., 2015). For instance, Savimäki et al. (2020) show that individuals who worried about becoming victims of online hate speech generally experienced more stress and were more likely to avoid stress-inducing situations.

The pragmatic vocabulary can be said to follow consequentialist reasoning that evaluates actions according to their presumed success in promoting desirable outcomes (Duffy & Freeman, 2011 in Hove, 2021, p. 888). Since success in promoting distant goals such as inclusive public discourse seems unlikely, in this vocabulary, intervention becomes an individual and optional activity, meaning that users are free to intervene if they want to or have the means to do so.

Discussion

People's actions are rarely motivated purely by their wish to be good citizens (Thorson, 2012, 2015). For this reason, we took a multifaceted approach to explore how users perceive and interpret norms of incivility and intervention and the implications of these perceptions for users' justification regarding their own and others' involvement in addressing incivility.

The presence of injunctive norms, representing social pressures to conform and avoid sanctions, theoretically implies a willingness to disapprove of deviant behavior (Chung & Rimal, 2016). However, most participants in this study approved of *non-intervention*. Therefore, without a personal inclination to employ even minor social sanctions, such as expressing disapproval, participants appear unlikely to intervene based on perceived injunctive norms. Instead, participants' considerations about intervention were more strongly shaped by their perceived descriptive norms. Indeed, research has suggested that descriptive norms of intervention may be comparably more effective in *preventing* incivility by signaling norms of "normal" behavior or informing the potential deviants that they can count on resistance, i.e., intervention of other users (Álvarez-Benjumea & Winter, 2018; Friess et al., 2021). Consistent with the Focus Theory of Normative Conduct, our findings indicate that the influence of descriptive norms may depend on the specific aspect of intervention users focus on. When users lack personal involvement, they seem to rely on agreeable descriptive norms (e.g., "incivility is everywhere," "intervention ends in conflict") to justify non-intervention, possibly to create emotional distance and minimize their role. The prominence of descriptive norms in our study may also be attributed to the broader context in which users encounter incivility, namely while habitually scrolling through their newsfeeds. When asked about their response to incivility in their newsfeeds, they commonly responded, "I keep scrolling." This aligns with research highlighting the influence of cognitive resources in activating injunctive and descriptive norms (Jacobson et al., 2011; Mark, 2023). During social media scrolling, individuals tend to be in an unfocused, shallow state of attention (Mark, 2023), making the activation of

descriptive norms more likely given their function as decision-making shortcuts (Jacobson et al., 2011). Indeed, continuing to scroll is much less costly and more personally gratifying than engaging in intervention to achieve more distant goals. However, these are only speculations. We encourage future research to consider the role of different attentional states in activating citizenship norms.

This tension between ideals and norms has been discussed previously (e.g., Gagrčin & Porten-Cheé, 2023; Thorson, 2012), and our study has tried to consolidate this relationship by looking at them jointly under the umbrella of normative vocabularies. We conceptualize vocabularies of intervention as specific ways of combining citizenship norms and ideals into coherent patterns of reasoning to justify (non-)intervention. We identify three distinct vocabularies: *aspirational, dismissive, and pragmatic*. The justifications primarily differ in how they perceived and weighted norms according to situations in vignettes and their personal experiences. The vocabularies illustrate different strategies for shifting responsibility and minimizing one's role in addressing incivility (Bandura, 2002), thereby alleviating dissonance with ideals of good citizenship one may hold. Specifically, dismissive and pragmatic justifications can be seen as *exit strategies* to avoid such dissonance. Finally, our findings suggest that the identified justifications are not solely properties of individuals; instead, the same person can employ all three in different situations (e.g., depending on their cognitive capacity, type of incivility, space where incivility occurs, etc.). We need more research to understand better the situational factors propelling different justifications.

Scholars have argued that a normative balance is (re) instated when enough people become norm followers and enforcers (Legros & Cislaghi, 2020), and norm salience plays a critical role in influencing behavior through norms. Nevertheless, regardless of the pervasiveness of a given norm, its potential to shape behavior remains improbable unless the norm is salient in the relevant moment. As mentioned earlier, user intervention online is commonly framed and promoted as a virtuous act of civil courage to achieve this salience. However, we speculate that contrary to activists (such as those in the #ichbinhier movement), who, as a community of practice, translate the ideals of discursive responsibility into concrete action on the group level (Passy & Monsch, 2020), individual users may have difficulty to justify the anticipated emotional cost of intervention and thus strongly rely on descriptive norms when deciding whether or not to intervene. In other words, promoting a sense of urgency by appealing to injunctive norms or ideals may not be as effective among those who experience fatigue and resignation.

On the practical level, our findings can help formulate appropriate empowerment and civic literacy strategies. Understanding the different vocabularies helps recognize the logic behind different justifications quickly. When engaging with dismissive justifications that tend to drift into cynicism,

it may be worthwhile to allow reflection on the relationship with social media and the possible personal and public consequences of a laissez-faire approach. Research shows that thinking about *reasons* for one's attitudes accentuates deep introspection and activates cognitive aspects relevant to actualizing injunctive norms (Bohner & Wänke, 2002, pp. 227-228). Individuals with a predominantly aspirational vocabulary could benefit from thinking about personal well-being while engaging in emotionally draining acts of intervention. Likewise, these individuals might benefit from reflecting on how their unapologetic idealism may be perceived as demotivating by non-interveners with legitimate reasons for inaction. Finally, reinstating normative balance seems to require *enough* users to intervene to establish intervention as a descriptive norm that one can realistically expect to actualize (Friess et al., 2021). Thus, individuals who employ the pragmatic line of reasoning, which we presume is most present (Emmer et al., 2021), would benefit from consciously developing heuristics to help them choose their battles and balance the normative demand for intervention and their capacities. An example of such heuristics could be assisting people in reflecting on a division of civic labor (Moe, 2020) and their social roles in different delineated spaces on social media platforms (Gagrčin, 2022). Specifically, there is a need to balance communicating user intervention as normatively desirable and practically feasible given the resources required and the readiness of social media users to invest them in everyday social media use.

Our typology is by no means exhaustive. First, normative vocabularies, as we examined in this study, depend on the social context. More research is needed to understand how older cohorts conceptualize their role in social media environments, especially since older generations tend to dwell on platforms such as Facebook, as young people increasingly move to platforms such as TikTok with different affordances, content, and norms (Literat & Kligler-Vilenchik, 2021). Similarly, it was beyond the possibilities of this study and, specifically, our sample to examine how gender, race, and ethnicity may shape norm perception and salience. Future research could examine how and to what extent these factors impact the construction and use of individual vocabularies. Quantitative studies could take an angle comparable to normative vocabularies, such as the media repertoires framework (Kim, 2016) or political toolkits (Oser, 2017). These could help assess online intervention in the context of other social media behaviors and normative perceptions, as well as specific patterns of intervention characteristics for user groups (Lane et al., 2022).

User intervention as a form of political engagement is less an enduring activity but an interruption of everyday life (Dahlgren, 2009)—a decision made while scrolling through social media. As a situated action, user intervention requires normative vocabularies that include individual awareness of the bigger societal context, acknowledgment of one's accountability, and readiness and skills to divert and invest their attention and resources relative to their possibilities. Is

it necessary to move more people toward an aspirational vocabulary? Or is it legitimate that not everyone wants to be as invested in enforcing norms on social media that most people use for escapist and entertainment purposes? This discussion goes beyond the scope of this paper and involves deeper scholarly reflection on the normative premises that inform academic inquiry and evaluation of users' intervening practices.

Declaration of Conflicting Interests

Funding

ORCID iD

Emilija Gagrčin   https://orcid.org/0000-0002-2953-1871

Note

1. In the German context, some initiatives include #ichbinhier and klicksafe.de. On the European level, the Council of Europe's campaign "No Hate Speech" has been running since 2013, seeking to combat hate speech through human rights education and awareness-raising, youth participation, and media literacy.

References

Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, *34*(3), 223–237. https://doi.org/10.1093/esr/jcy005

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, *31*(2), 101–119. https://doi.org/10.1080/0305724022014322

Bohner, G., & Wänke, M. (2002). *Attitudes and attitude change*. Psychology Press.

Bormann, M., Tranow, U., Vowe, G., & Ziegele, M. (2022). Incivility as a violation of communication norms—A typology based on normative expectations toward political communication. *Communication Theory*, *32*(3), 332–362. https://doi.org/10.1093/ct/qtab018

Buerger, C. (2021). #iamhere: Collective counterspeech and the quest to improve online discourse. *Social Media + Society*, *7*(4). https://doi.org/10.1177/20563051211063843

Chung, A., & Rimal, R. N. (2016). Social norms: A review. *Review of Communication Research*, *4*, 1–28. https://doi.org/10.12840/issn.2255-4165.2016.04.01.008

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, *24*(C), 201–234. https://doi.org/10.1016/S0065-2601(08)60330-5

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. https://doi.org/10.1037/0022-3514.58.6.1015

Dahlgren, P. (2009). *Media and political engagement: Citizens, communication, and democracy*. Cambridge University Press.

Dalton, R. J. (2008). Citizenship norms and the expansion of political participation. *Political Studies*, *56*(1), 76–98. https://doi.org/10.1111/j.1467-9248.2007.00718.x

Emmer, M., Leißner, L., Strippel, C., Porten-Cheé, P., & Schaetz, N. (2021). *Weizenbaum Report 2021: Politische Partizipation in Deutschland* [Weizenbaum Report 2021: Political participation in Germany]. https://doi.org/10.34669/wi.wr/2

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Finch, J. (1987). The vignette in technique in survey research. *Sociology*, *21*(1), 105–114.

Fowler, J. H., & Kam, C. D. (2003). Beyond the self: Social identity, altruism, and political participation. *Journal of Politics*, *69*(3), 811–825.

Friess, D., Ziegele, M., & Heinbach, D. (2021). Collective civic moderation for deliberation? Exploring the links between citizens' organized engagement in comment sections and the deliberative quality of online discussions. *Political Communication*, *38*(5), 624–646. https://doi.org/10.1080/10584609.2020.1830322

Gagrčin, E. (2022). Your social ties, your personal public sphere, your responsibility: How users construe a sense of personal responsibility for intervention against uncivil comments on Facebook. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448221117499

Gagrčin, E., & Porten-Cheé, P. (2023). Between individual and collective social effort: Vocabularies of informed citizenship in different information environments. *International Journal of Communication*, *17*, 1510–1529. https://ijoc.org/index.php/ijoc/article/view/17240/4065

Gagrčin, E., Porten-Cheé, P., Leißner, L., Emmer, M., & Jørring, L. (2022). What makes a good citizen online? The emergence of discursive citizenship norms in social media environments. *Social Media + Society*, *8*(1). https://doi.org/10.1177/20563051221084297

Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, *7*(2). https://doi.org/10.1177/2053951720943234

Gray, D., Royall, B., & Malson, H. (2017). Hypothetically speaking: Using vignettes as a stand-alone qualitative method. In V. Braun, V. Clarke, & D. Gray (Eds.), *Collecting qualitative data: A practical guide to textual, media and virtual techniques* (pp. 71–93). Cambridge University Press.

Heger, K., Leißner, L., Emmer, M., & Strippel, C. (2022). *Weizenbaum Report 2022: Politische Partizipation in Deutschland* [Weizenbaum Report 2022: Political participation in Germany]. https://doi.org/10.34669/wi.wr/3

Hove, T. (2021). Media ethics, moral controversies, and the sociology of critique. *Communication Theory*, *31*(4), 884–904. https://doi.org/10.1093/ct/qtaa016

Jacobson, R. P., Mortensen, C. R., & Cialdini, R. B. (2011). Bodies obliged and unbound: Differentiated response tendencies for injunctive and descriptive social norms. *Journal of Personality and Social Psychology*, *100*(3), 433–448. https://doi.org/10.1037/a0021470

Kalch, A., & Naab, T. K. (2017). Replying, disliking, flagging: How users engage with uncivil and impolite comments on news sites. *Studies in Communication—Media*, *6*(4), 395–419. https://doi.org/10.5771/2192-4007-2017-4-395

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, *26*(8), 1002–1012. https://doi.org/10.1177/01461672002610009

Kim, S. J. (2016). A repertoire approach to cross-platform media use behavior. *New Media & Society*, *18*(3), 353–372. https://doi.org/10.1177/1461444814543162

Kligler-Vilenchik, N. (2017). Alternative citizenship models: Contextualizing new media and the new "good citizen." *New Media & Society*, *19*(11), 1887–1903. https://doi.org/10.1177/1461444817713742

Kunst, M., Porten-Cheé, P., Emmer, M., & Eilders, C. (2021). Do "good citizens" fight hate speech online? Effects of solidarity citizenship norms on user responses to hate comments. *Journal of Information Technology & Politics*, *18*(3), 258–273. https://doi.org/10.1080/19331681.2020.1871149

Kwak, N., Lane, D. S., Weeks, B. E., Kim, D. H., Lee, S. S., & Bachleda, S. (2018). Perceptions of social media for politics: Testing the slacktivism hypothesis. *Human Communication Research*, *44*(2), 197–221. https://doi.org/10.1093/hcr/hqx008

Lane, D. S., Weeks, B. E., & Kwak, N. (2022). A repertoire approach to studying inequalities in political social media use. *Mass Communication and Society*. Advance online publication. https://doi.org/10.1080/15205436.2022.2137039

Lapinski, M. K., & Rimal, R. N. (2005). An explication of social norms. *Communication Theory*, *15*(2), 127–147. https://doi.org/10.1093/ct/15.2.127

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Englewood Cliffs.

Legros, S., & Cislaghi, B. (2020). Mapping the social-norms literature: An overview of reviews. *Perspectives on Psychological Science*, *15*(1), 62–80. https://doi.org/10.1177/1745691619866455

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—Media*, *7*(4), 555–579. https://doi.org/10.5771/2192-4007-2018-4-555

Lindenberg, S. M. (2008). Social norms: What happens when they become more abstract? In A. Diekmann, K. Eichner, P. Schmidt, & T. Voss (Eds.), *Rational Choice: Theoretische Analysen und empirische Resultate* (pp. 63–81). VS Verlag für Sozialwissenschaften.

Literat, I., & Kligler-Vilenchik, N. (2021). How popular culture prompts youth collective political expression and cross-cutting political talk on social media: A cross-platform analysis. *Social Media + Society*, *7*(2). https://doi.org/10.1177/20563051211008821

Mark, G. (2023). *Attention span*. Hanover Square Press.

Miškolci, J., Kováčová, L., & Rigová, E. (2020). Countering hate speech on Facebook: The case of the Roma minority in Slovakia. *Social Science Computer Review*, *38*(2), 128–146. https://doi.org/10.1177/0894439318791786

Moe, H. (2020). Distributed readiness citizenship: A realistic, normative concept for citizens' public connection. *Communication Theory*, *30*(2), 205–225. https://doi.org/10.1093/ct/qtz016

Naab, T. K. (2016). Der Sanktionsbedarf von Facebook-Inhalten aus Sicht von NutzerInnen und seine Determinanten [The need for sanctioning Facebook content from the perspective of users and its determinants]. *Medien & Kommunikationswissenschaft*, *64*(1), 56–73. https://doi.org/10.5771/1615-634X-2016-1-56

Naab, T. K., Kalch, A., & Meitz, T. G. K. (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. https://doi.org/10.1177/1461444816670923

Naab, T. K., Naab, T., & Brandmeier, J. (2021). Uncivil user comments increase users' intention to engage in corrective actions and their support for authoritative restrictive actions. *Journalism & Mass Communication Quarterly*, *98*(2), 566–588. https://doi.org/10.1177/1077699019886586

Neubaum, G., Cargnino, M., Winter, S., & Dvir-Gvirsman, S. (2021). "You're still worth it": The moral and relational context of politically motivated unfriending decisions in online networks. *PLOS ONE*, *16*(1). https://doi.org/10.1371/journal.pone.0243049

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Reuters Institute Digital News Report 2022*. Reuters Institute.

Oser, J. (2017). Assessing how participators combine acts in their "political tool kits": A person-centered measurement approach for analyzing citizen participation. *Social Indicators Research*, *133*(1), 235–258. https://doi.org/10.1007/s11205-016-1364-8

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Passy, F., & Monsch, G.-A. (2020). *Contentious minds. How talk and ties sustain activism*. Oxford University Press.

Porten-Cheé, 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*(2020), 514–534.

Ragin, C. C., & Amoroso, L. M. (2011). *Constructing social research*. Pine Forge Press.

Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Räsänen, P., & Zych, I. (2021). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior*, *42*(9), 1100–1111. https://doi.org/10.1080/01639625.2020.1722337

Rimal, R. N., & Real, K. (2003). Understanding the influence of perceived norms on behaviors. *Communication Theory*, *13*(2), 184–203. https://doi.org/10.1111/j.1468-2885.2003.tb00288.x

Robison, J. (2022). Valuing politics: Explaining citizen's normative conceptions of citizenship. *Political Behavior*. Advance online publication. https://doi.org/10.1007/s11109-022-09773-7

Rossini, P. (2022). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk.

Communication Research, *49*(3), 399–425. https://doi.org/10.1177/0093650220921314

Savimäki, T., Kaakinen, M., Räsänen, P., & Oksanen, A. (2020). Disquieted by online hate: Negative experiences of Finnish adolescents and young adults. *European Journal on Criminal Policy and Research*, *26*(1), 23–37. https://doi.org/10.1007/s10610-018-9393-2

Shmargad, Y., Coe, K., Kenski, K., & Rains, S. A. (2022). Social norms and the dynamics of online incivility. *Social Science Computer Review*, *40*(3), 717–735. https://doi.org/10.1177/0894439320985527

Stok, F. M., & de Ridder, D. T. D. (2019). The focus theory of normative conduct. In K. Sassenberg & M. L. W. Vliek (Eds.), *Social psychology in action* (pp. 95–110). Springer.

Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review*, *51*(2), 273–286.

Thorson, K. (2012). What does it mean to be a good citizen? Citizenship vocabularies as resources for action. *Annals of the American Academy of Political and Social Science*, *644*(1), 70–85. https://doi.org/10.1177/0002716212453264

Thorson, K. (2015). Sampling from the civic buffet: Youth, new media and do-it-yourself citizenship. In H. Gil de Zúñiga (Ed.), *New technologies & civic engagement: New agendas in communication* (pp. 3–22). Routledge.

Tworek, H., & Leerssen, P. (2019). *An analysis of Germany's NetzDG Law*. Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression. https://hdl.handle.net/11245.1/3dc07e3e-a988-4f61-bb8c-388d903504a7

Vraga, E. K., Thorson, K., Kligler-Vilenchik, N., & Gee, E. (2015). How individual sensitivities to disagreement shape youth political expression on Facebook. *Computers in Human Behavior*, *45*, 281–289. https://doi.org/10.1016/j.chb.2014.12.025

Willems, J. (2021). Learning civil courage: A citizens' perspective. *Educational Researcher*, *50*(9), 679–681. https://doi.org/10.3102/0013189X211044159

Ziegele, M., Naab, T. K., & Jost, P. (2020). Lonely together? Identifying the determinants of collective corrective action against uncivil comments. *New Media & Society*, *22*(5), 731–751. https://doi.org/10.1177/1461444819870130

Author Biographies

Emilija Gagrčin (PhD, Freie Universität Berlin) is a researcher and lecturer at the Institute for Media and Communication Studies at the University of Mannheim and an associated researcher in the Department of Information Science and Media Studies at the University of Bergen. She is affiliated with the Weizenbaum-Institute for the Networked Society. Her research interests include social, normative, and infrastructural aspects of platformized civic life.

Miriam Milzner (MA, Freie Universität Berlin) is a research associate and doctoral candidate at Freie Universität Berlin and the Weizenbaum-Institute for the Networked Society. Drawing from communication science and computational social science, she analyzes the logics of digital infrastructures, the hybrid media sphere, and the dynamics of (automated and coordinated) information manipulation.