Article

new media & society
2024, Vol. 26(8) 4299–4316

DOI: 10.1177/14614448221117499
journals.sagepub.com/home/nms


Your social ties, your personal public sphere, your responsibility: How users construe a sense of personal responsibility for intervention against uncivil comments on Facebook

Emilija Gagrčin
Freie Universität Berlin, Germany

Abstract

User intervention against incivility is a significant element of democratic norm enforcement on social media, and feeling personally responsible for acting is a vital prerequisite for intervention. However, our insight into how users construe their sense of personal responsibility and expectations of other users remains limited. By theoretically foregrounding user perspective, this study investigates the boundaries and nuances of user responsibility to intervene against incivility. Empirically, it draws on 20 qualitative vignette interviews with young people in Germany. The findings show that as contexts collapse in users' newsfeeds, the imagined boundaries of personal public spheres and own social relationships with uncivil users serve as heuristics for hierarchizing and delimiting personal responsibility to intervene. Beyond abstract individual responsibility for the public discourse, practical responsibility is distributed among personal public spheres.

Keywords

Bystander intervention, hate speech, incivility, user comments, user intervention

Corresponding author:
Emilija Gagrčin, Institut für Publizistik- und Kommunikationswissenschaft, Freie Universität Berlin, Garystr. 55, 14165 Berlin, Germany.
Email: emilija.gagrcin@fu-berlin.de

Uncivil discourse online is a growing concern among citizens and scholars alike, as it pollutes the public discourse and has exclusionary implications for minority participation (Anderson et al., 2014; Porten-Cheé et al., 2020; Ziegele et al., 2020). Of the numerous platforms available, users are most likely to encounter hateful content on Facebook (Reichelmann et al., 2021). Although Facebook recently introduced measures such as automated content moderation (Meta, 2021, 2022), technological solutions for countering incivility and hate fall short when contextual interpretation is required (e.g. Gillespie, 2020; Meta, 2022; Siapera and Viejo-Otero, 2021). Given the amount of problematic content that remains on the platform (Giansiracusa, 2021; Timberg, 2021), ordinary users as co-constructors of social media environments remain relevant for restoring favorable conditions for political discourse (Friess et al., 2020; Kim, 2021; Masullo et al., 2019; Meta, 2022; Porten-Cheé et al., 2020; Ziegele et al., 2020).

A sense of personal responsibility is vital in prompting individuals to intervene against incivility (Latané and Darley, 1970; Ziegele et al., 2020). This has been illustrated by research into the social movement #iamhere, in which users are motivated to engage by a sense of responsibility for the public discourse (Ziegele et al., 2020). Other studies have indicated that some users intervene out of solidarity (Kunst et al., 2021) or for altruistic reasons (Wang and Kim, 2020). However, user intervention against incivility is overall not that common. For example, repeated representative surveys in Germany have shown that while most people believe standing up to discrimination and hate speech to be a sign of good citizenship, only a minority report intervening upon encountering these online (Emmer et al., 2021; Heger et al., 2022; Schaetz et al., 2020). This suggests that regular users either do not feel a concrete sense of personal responsibility to act or have a different understanding of responsibility altogether. Nevertheless, despite studies demonstrating the pivotal role of personal responsibility for intervention (Latané and Darley, 1970; Ziegele et al., 2020), our understanding of how regular users make sense of their role in combating incivility is surprisingly limited. I argue that it is also obscured by our normative approach to studying user intervention, which is grounded in scholarly imaginaries of the online public sphere and a perspective on individual action as decoupled from the social context in which it occurs (cf. Dahlgren, 2006).

Since users experience public discourse on Facebook through their news feeds, where different public, private, political, and social contexts converge (Marwick and boyd, 2011), I theorize that understanding users' perceptions of responsibility requires considering other everyday social media experiences known to shape sociability and informal political talk. Drawing on the literature on online boundary, relational and impression management, I investigate how Facebook users construe a sense of personal responsibility to intervene against incivility in the context of their everyday social media use. Based on 20 vignette interviews with students in Germany, this study shows that users feel most strongly compelled to intervene when incivility occurs in what they perceive as their *personal public sphere*—a delineated communicative space of their own that intersects with and is visible to others, which creates the need for impression management and a sense of personal accountability (John and Gal, 2018). In this space, an intervention is considered comparably more meaningful and

efficacious because it involves significant social ties, as opposed to intervening in news media comment sections that involve unknown users. Thus, the boundaries of one's personal public sphere and social relatedness to uncivil users serve as heuristics for thinking about the practical and immediate responsibility to intervene. The findings remind us that not everything that we (both as scholars and social media users) deem normatively desirable is practically feasible, appropriate, or immediately important in the context of everyday social media use. This study contributes to a further understanding of discursive civic responsibility by offering a perspective on responsibility as distributed among proprietors of *personal* public spheres rather than as diffused among individual bystanders in *the* public sphere.

Literature review

Incivility and intervention in user comments

Social media platforms afford different opportunities for political talk and self-expression (Literat and Kligler-Vilenchik, 2021). On Facebook, in particular, informal political talk features prominently in user comments. Scholars have argued in support of comments' democratic benefits (Freelon, 2015), even when they do not conform to standards of rationality and politeness (Rossini, 2022). In contrast, an increasing amount of *uncivil* discourse in user comments has been seen as troubling (Anderson et al., 2014; Reichelmann et al., 2021). As a mode of expression that "signals moral disrespect and profound disregard toward individuals or groups" (Rossini, 2022: 7), incivility can be viewed as disregarding democratic values such as pluralism (Rossini, 2022) and violating moral (Neubaum et al., 2021b) and/or communicative norms (Bormann et al., 2021). As a counter-normative behavior, incivility is likely to attract condemnation, censure, and punishment by relevant audiences (Watson et al., 2019).

Platforms offer different modalities for sanctioning incivility. Besides reporting uncivil content to Facebook as a violation of the platform's Community Guidelines (Meta, 2022; Siapera and Viejo-Otero, 2021), users can voice their disapproval by reacting to uncivil comments with angry emojis or through counter-commenting (Masullo and Kim, 2021; Porten-Cheé et al., 2020). By engaging in such interventions, people are said to "seek to voice their own opinions to correct the 'wrongs' they perceive in the public sphere" (Barnidge and Rojas, 2014: 136) and aim to "ensure an inclusive online public discourse" (Porten-Cheé et al., 2020: 519). In this context, the bystander intervention model postulates that feeling a sense of *personal* responsibility to act is a vital prerequisite for intervention, followed by a decision on how to intervene *appropriately* (Latané and Darley, 1970; Ziegele et al., 2020). A prominent explanation for user inaction is the so-called bystander effect, according to which the presence of others leads to a diffusion of responsibility and results in a disinclination to act (Latané and Darley, 1970).

To date, the most nuanced insights into users' ideas of personal responsibility stem from research into the social movement #iamhere, in the context of which users engage in collective intervention to promote a cultivated discourse on Facebook (Buerger, 2021; Friess et al., 2020; Ziegele et al., 2020). These users report being motivated by a sense of

personal responsibility for the public discourse (Ziegele et al., 2020). Most people in Germany approve of this kind of engagement. Repeated representative surveys show that more than 70% of respondents believe that standing up to hate and discrimination is good citizenship (Emmer et al., 2021; Schaetz et al., 2020). However, only a minority of the survey respondents report actually having intervened upon encountering incivility (Emmer et al., 2021; Heger et al., 2022), underscoring that activists and non-activists differ in their mindsets and in their abilities to sustain a sense of responsibility and motivation for (collective) action (Passy and Monsch, 2020).

From a normative point of view, #iamhere's engagement appeals to some of the central premises of research into bystander intervention against incivility: (1) news media comment sections are central spaces for public deliberation online, (2) users act in their role as citizens, and (3) users are equals in social media environments, entitled to sanction each other based on their horizontal relationship as citizens (Dishon and Ben-Porath, 2018). However, there are good reasons to believe that other aspects of mediated social life on Facebook inform users' ideas about responsibility for intervention. Extant literature suggests that users' perceptions of self and others shape not only the choreography of social interactions online (Baym and boyd, 2012; Kligler-Vilenchik, 2021; Literat and Kligler-Vilenchik, 2021; McLaughlin and Vitak, 2012; Marwick and boyd, 2011) but also people's notions of citizenship (Gagrčin et al., 2022; Gagrčin and Porten-Cheé, 2022). In the following, I consider how spatial and social aspects of everyday social media experience may shape regular users' sense of personal responsibility to intervene against incivility.

Public sphere(s)

Studies have typically focused on incivility in user comments on news organizations' websites and social media pages due to their attributed function as deliberative public spaces (e.g. Freelon, 2015; Kim, 2021; Stroud et al., 2015; Watson et al., 2019). However, this does not necessarily resonate with how users imagine and navigate social media environments. More than a decade ago, Papacharissi (2010) argued that social media would blur the boundaries between public and private spaces in a way that alters "the actual and imagined spaces upon which citizenship is practiced" (p. 17). Studies have demonstrated that users are more likely to interact with their Friends' news posts than posts on news pages (Wells and Thorson, 2017), which challenges the notion of *the* public sphere where political talk online occurs and, relatedly, user responsibility to intervene out of responsibility for *the* public discourse. Moreover, John and Gal (2018) have found that users do not necessarily imagine or experience Facebook as one big public sphere but rather as a more delineated *personal* public sphere—a communicative space of their own with specific boundaries. How users visualize these boundaries presumably differs between platforms (Literat and Kligler-Vilenchik, 2021). On Facebook, the personal public sphere can include users' profiles, news feed, and friends' lists. Aware that their personal public sphere intersects with others' personal public spheres, users believe they have both the right and the obligation to regulate and curate content and interactions based on their own norms and values (John and Agbarya, 2021; John and Gal, 2018; Schmidt, 2014).

Face-work and relational work

The backbone of the personal public sphere concept is the centrality of face-work and relational work in the context of mediated social life (John and Gal, 2018; Schwarz and Shani, 2016). In everyday social interactions, individuals engage in face-work by acting according to their perceptions of audience expectations to maintain "the positive social value they claim for themselves" (Goffman, 1967: 5). Face-work is particularly laborious on Facebook. As different spheres of life converge, face-work is done before multiple audiences simultaneously: close ties, such as friends and family members, and more distant ties, such as acquaintances from school or friends of friends (Baym and boyd, 2012; Marwick and boyd, 2011; Schwarz and Shani, 2016). At the same time, however, the audience on Facebook is not visible to users. Instead, how users *imagine* their audience is crucial to their situational public self-awareness and perceptions of behavioral expectations (Litt, 2012; Mor et al., 2015). Here, users' *civic* role is but one of many social roles that users may assume when thinking about what is expected of them.

A "mismanagement" of the online self may have real-life consequences—particularly for interpersonal relationships (John and Agbarya, 2021; Mor et al., 2015). Faced with an uncertain reception, some users preemptively engage in self-censorship or abstain from political talk and self-expression to avoid conflict and mitigate risks (Pearce et al., 2018; Thorson, 2014; Vraga et al., 2015). Others yet unfriend social ties for posting problematic content—either because they do not want to see that kind of content anymore or because they do not want to be associated with those users (Gagrčin et al., 2022; John and Agbarya, 2021; John and Gal, 2018).

Interventions such as counter-speaking are arguably more confrontational than political unfriending and can be seen as socially delicate endeavors. In contrast to the idea that users are entitled to sanction each other based on their horizontal relationship as citizens (Dishon and Ben-Porath, 2018), social relationships shape perceptions of who is responsible for intervening (Moisuc and Brauer, 2019; Strimling and Eriksson, 2014). Research shows that in the presence of friends and strangers, friends—not strangers—are expected to sanction (Eriksson et al., 2017; Strimling and Eriksson, 2014). At the same time, our relationships influence how we judge and react to norm violations (McLaughlin and Vitak, 2012), and people tend to be harsher toward distant ties as opposed to close ones (Lieberman and Linke, 2007; Neubaum et al., 2021b). Thus, the need for face-work and relational work may motivate and/or constrain one's sense of responsibility to act against incivility.

To better understand how personal responsibility compels users to counter incivility, I conceptualize responsibility not only in terms of desirability (what a good citizen would do) but also in terms of behavioral expectations that individuals perceive and place upon themselves and others (Cialdini et al., 1991), and ask the following questions:

What is the role of personal public sphere(s) (RQ1), impression management (RQ2), and social relatedness (RQ3) in people's expectations of user intervention against incivility on social media?

Methods

Study context

In contrast to the United States, where the First Amendment to its constitution guarantees freedom of speech, the German Constitution perpetuates the idea 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" (Riedl et al., 2021: 437). In addition, Germany's Network Enforcing Act provides users with critical agency in social media environments by requiring platforms to delete problematic content, for example, when flagged by users (Heldt, 2019). Perhaps unsurprisingly, a recent study has indicated that Germans have comparably higher expectations of governmental regulation of hate speech and incivility online than Americans and assume comparably higher levels of personal responsibility for intervention (Riedl et al., 2021). Thus, Germany provides an ideal context for exploring how users practically construe this responsibility.

Facebook is a relevant case for several reasons. In addition to introducing artificial intelligence to detect hate speech, Facebook still relies on the idea of self-regulation, expecting users to proactively report content that they believe violates the Community Guidelines (Meta, 2022). After the reported success of automated hate speech detection was repeatedly called into question (most recently by whistleblower Frances Haugen), proactive norm enforcement on the part of users has become particularly important (Giansiracusa, 2021; Timberg, 2021). Finally, because users are more likely to encounter hateful content on Facebook than on other platforms (Reichelmann et al., 2021), users may feel that Facebook is a space in particular need of user intervention.

Participants

The study draws upon 20 semi-structured interviews with German university students, ages 20–25 years. The decision to study this sample was based on the following considerations: First, young people use social media for political purposes more commonly than older adults (Andersen et al., 2020; Emmer et al., 2021). Second, studies have shown that younger and wealthier people are more likely to intervene (Watson et al., 2019). The participants were recruited via university email lists, where they registered and filled out a pre-screening questionnaire. The questionnaire aimed to recruit a diverse sample of respondents and avoid intervention enthusiasts' self-selection bias. In the final sample, most participants self-identified as rare interveners; only three identified as occasional interveners. The average age is 22 years, with 40% of participants self-identifying as male and 60% as female. All participants used at least two social media platforms daily and had Facebook profiles. I use pseudonyms chosen by the participants to report on the study, and I have translated the quotes used into English.

Interviews

Vignette interviews were employed as the standalone method in this study because the method is suitable for constructivist approaches that explore participants' ethical frameworks and moral codes (Gray et al., 2017; Wilks, 2004). The participants were presented

with two fictional Facebook posts with accompanying texts (see Supplemental Appendix). The first depicted incivility in the comment section below a news post on a user profile (representing personal space on Facebook). The second instance of incivility was situated in the comment section below a news post on a German news outlet's page (representing public space). Based on the literature showing that people recognize impoliteness much more easily than incivility (Kalch and Naab, 2017), both uncivil comments were formulated as polite. Because I was interested in how people define their responsibility to intervene, I needed to ensure that participants perceived the comments as uncivil—a step that precedes defining responsibility in the bystander intervention model (Ziegele et al., 2020). As previous research has indicated that abusive language directed at social groups is considered particularly threatening (Naab et al., 2018; Wilhelm et al., 2020), I chose refugees and people with disabilities as targets of incivility, assuming most participants would likely condemn discrimination against these two groups. The vignettes were tested in five trial interviews to ensure that the situations appeared typical and realistic; following participant feedback, these were further adjusted.

After reading the vignettes, the participants assessed the situation, after which they were asked to take on several roles in different relationship constellations (Gray et al., 2017; O'Dell et al., 2012). For example, I asked participants what they believed the post owner ought to do in a situation in which the uncivil user was their friend and whether it would make a difference if they were an acquaintance from school or a stranger. I encouraged participants to reminisce and reflect on similar situations that had happened to them or that they had observed. Although I had concerns about whether the participants would be able to switch from one role to another, it was surprisingly effortless for most of them.

The interviews were conducted via video conferencing platforms and lasted approximately 80 minutes. Only audio was recorded. Student assistants transcribed the interviews, and I coded them. The analysis was conducted according to Saldaña (2016). In the first step, I exploratively coded a subset of interviews ($n = 5$) using in-vivo and versus coding to develop the initial codes list. Because I was interested in responsibility not only in terms of desirability but also in terms of social expectations (Cialdini et al., 1991), I coded the former as actions that participants *wished would* happen, would *ideally* happen, or were *theoretically important*, and the latter in terms of *must, should*, and *ought to do*. In the second stage, I consolidated the codes according to the roles, rules, and relationships in Saldaña and Omasta (2018) and applied them to the rest of the interviews.

Findings

Similar to the ideas about responsibility reported and conceptualized in the literature, participants recognized incivility as problematic and worrisome. Most shared the view that responding to incivility is, in principle, a civic responsibility, corroborating the findings of other interview studies (Ziegele et al., 2020). However, participants stressed that this was, first and foremost, an abstract responsibility—something that one would *ideally* do—adding that there were many limitations and good reasons *not to act* upon this responsibility in practice. For example, participants believed the vignettes were likely to

produce conflict. Franziska (25 years) was certain, "It's about to get a lot more unpleasant . . . someone will feel attacked, especially if they know each other." In this sense, participants frequently emphasized that intervention requires a great deal of time and emotional resources. However, of greater interest here are the instances in which participants felt that intervention was, in fact, a matter of personal responsibility.

Personal public sphere and the responsibility to intervene in public

Participants placed the strongest expectations for intervention against incivility on the profile or page proprietor on whose *territory* incivility occurred. They spoke in terms that concurred with the concept of the personal public sphere and the idea that one has both authority over, and obligations to, others in that delineated communicative space (John and Gal, 2018). Two quotes from participants neatly encapsulate this idea:

> It's simply how Facebook works—it's your account, so whatever you post, it's your platform. And everyone who sees your post in their news feed is exposed to it. So, I think you are responsible for trying to keep your page free of discrimination. (Naomi, 22 years)

> If you post something about people with disabilities, like in this example, you are also taking on a role to speak for them and their rights. And if someone denigrates them, then I think you should stay on the ball and be able to defend this group and essentially your positions. (Henri, 20 years)

Both illustrate that the desire to maintain a positive image of oneself creates an expectation that one would and should defend and enforce one's values and positions (John and Gal, 2018). Moreover, the perceptions of the personal public sphere reveal that user intervention is infused with several meanings. As Naomi articulated, the expectation is that users *publicly* signal to their audience that uncivil behavior is not tolerated in their public sphere. This signaling aims to show solidarity with the discriminated group, motivated by the idea of preventing the presumed influence of discriminatory content on the audience (Wang and Kim, 2020; Wintterlin et al., 2021). Participants considered intervention a form of social sanctioning that informs the uncivil user "that it's not okay to spread hate and lies" (Mark, 21 years) so that the uncivil user "experiences public pushback and maybe even realizes that what they said is wrong" (Charlotte, 25 years).

Despite having asserted that one should not leave incivility in one's personal public sphere unanswered, participants generally bemoaned the hollowness of such interventions. They often complained, "It's not even a real discussion but a stringing together of statements, where people reduce each other to these single short sentences" (Rebeca, 21 years). Sharing the same sentiment, many participants described how best to avoid a long discussion upon intervening publicly:

> The problem is that once you comment, it goes back and forth forever [laughs]. And other people interfere as well. And that's why I think it's important to take time to formulate a response so as not to offer much room for further discussion, umm, so that it doesn't drag on and get worse. (Franziska, 25 years)

These responses indicate that participants were generally not interested in seeking conversation with uncivil users—at least not in the comment section. Instead, they were intervening "for the record"—so that "in case someone stumbles upon the post, [the uncivil comment is] not the only comment they see" (Rebeca). By intervening, users consciously create artifacts for the judgment of audiences, evoking the view of social interaction on social media as an exhibition rather than performance (Hogan, 2010).[1]

The underlying image of the public discourse as an exhibition of fragments from different personal public spheres that constantly flow in and out of our news feed (Marwick and boyd, 2011; Thorson and Wells, 2016) creates a sense of personal responsibility to combat incivility in the *personal* public sphere. As a means of impression management, participants felt the urge to intervene because an absence of intervention was seen as an artifact testifying to users' failure to take care of their personal public sphere and stand up for themselves. They also acknowledged an obligation to other users: recognizing that fragments of their personal public sphere appeared in others' newsfeed compelled them to enforce discursive norms and ensure a certain quality in their part of the public discourse (Gagrčin et al., 2022; John and Gal, 2018).

Social relations and responsibility to reform in private

Social relationships between vignette characters influenced how interviewees read the vignettes and how they formulated the need and appropriateness of intervention. As Mark poignantly stated, with more distant social ties, "[I]t's so easy to reduce their whole life to this one post and to think that they are idiots or Nazis. But when you've known people, you want to know how they came to think this way." Because friends are extensions of the self, we generally expect similarity and reciprocity from them (e.g. Hall, 2012). The closer the uncivil user was to the owner of the personal public sphere where incivility occurred, the more likely participants were to read the situation as an issue of disagreement and ground their normative irritation in the difference of opinions and the public display of this difference. Observing a situation in which a friend acts counternormatively produced a sense of cognitive dissonance, which people strove to mitigate by reinterpreting the situation (Festinger, 1957). Consider how May (24 years) read the situation and negotiated the need for intervention:

> I think if it's an entirely unknown person or just, I don't know, a former acquaintance from school, then you can just delete it and forget it. Now, I'd feel deep disappointment if it's a good friend. I would be like, "Oh wow, am I friends with the wrong person?" or "Is this person having a bad day?" So many negative feelings come up . . . You'd rather teach them or at least try to understand them. With strangers or people who have become strangers to me, I wouldn't give a damn . . . I would simply delete the comment and forget the person. But you want to get rid of the negative emotions you suddenly have for a friend.

Like May, participants typically emphasized the urge to *reform* the uncivil friend and believed they had an educational task (Hofmann et al., 2018; Neubaum et al., 2021a). Owing to the common ground they share with their close social ties, participants felt that the legitimacy and influence of their intervention might be comparably more significant to that of strangers.

Social proximity with an uncivil user shapes not only the meaning but also the appropriateness of intervention. When directed at distant ties, intervention essentially sanctions uncivil behavior (Tsfati and Dvir-Gvirsman, 2018). Making the uncivil user uncomfortable is arguably one of the goals of the pushback. In contrast, participants were wary of the face threat that a public pushback could cause to their close ties. Thus, participants considered it more appropriate to manage public interventions "backstage" by talking to uncivil users privately, asking them to remove their uncivil comments, or informing them that they would remove the comments themselves. Mark described this rationale playfully:

> If I knew [the uncivil commentator], I would definitely first seek a private conversation rather than exposing him so publicly! At the same time, he wrote [the uncivil comment] deliberately. He is accountable for it. But it's like seeing your pal step in dog poop in public, and instead of just going to the person and quietly offering them tissues, you start yelling "Watch out, dog poop!" and pointing fingers at the person: "Look, he stepped in dog poop!" That doesn't really help the cause.

Mark's input also reveals a reinterpretation of the situation by framing the uncivil comment as a disagreeable "incident" that can be overcome if one reacts appropriately. Beyond being a function of relational and impression management, retreating backstage can be seen as an intervention strategy. To have a chance at reforming an uncivil user, the participants believed they must limit the scope of the audience. Lola (20 years) explained, "[I]f you know [the uncivil user] and you wanted to talk them out of their point of view, you should try to speak to them privately. If you do it publicly, people react with fright or act dismissively." Taking a conversation backstage allows for a more intimate atmosphere where "both can display emotions and insecurities instead of demonizing each other" (Friedrich, 25 years). In this sense, while public intervention prompts artifact creation, backstage interventions seek to bypass the exhibition character by "re-insert[ing] situational definition into the technically converged experience of political talk" (Papacharissi, 2010: 73).

Boundaries of the personal public sphere and displacement of responsibility

Finally, perceptions of the personal public sphere inform how people think about personal responsibility and the appropriateness of bystander intervention in others' personal public spheres. Despite a shared belief that bystanders who care for the topic or the group addressed by the uncivil comment would be *inclined* to intervene (Kunst et al., 2021; Naab et al., 2018), participants did not *expect* bystanders to intervene, nor did they express a sense of personal responsibility regarding their intervention when assuming a bystander role. Echoing Naomi's and Henri's input from the beginning of the section, participants' sense of immediate personal responsibility and their expectations toward other users were most pronounced within the boundaries of the personal public sphere and decreased with the perceived social distance from the uncivil user and users whose personal public spheres had been affected.

Research has shown that a fear of embarrassment and being negatively judged by other bystanders hinders intervention (Kim, 2021; Van Bommel et al., 2012). In this study, however, the participants were not worried about other bystanders. Instead, they focused on the proprietor of the personal public sphere that had been affected, expressing a great deal of relational discomfort with meddling in their personal public sphere. This was particularly pronounced in relation to "unnuanced" social ties (Donath, 2007), and participants were hesitant to get involved without knowing the relationship between the users involved in the uncivil incident. One could easily dismiss an assertion such as "I wouldn't necessarily want to interfere in their relationship" (Timo, 25 years) as a mere excuse for non-intervening. However, as a recurring explanation for non-intervention, it indicates that user intervention as a social sanction is itself subject to norms, where social proximity functions as a heuristic for construing a sense of responsibility and appropriateness to sanction misbehavior (Moisuc and Brauer, 2019; Strimling and Eriksson, 2014).

Following the logic of a delimited space of responsibility for norm enforcement, participants did not feel responsible for intervening against uncivil comments below news media outlets' Facebook posts when stumbling upon them in their news feed. Instead, they expected these pages to allocate sufficient resources to comment moderation and strongly disapproved of their failure to intervene:

> I definitely have a different expectation [of news media pages] than private people. I mean, they are news providers! They are regularly confronted with [incivility], and they should have a strategy for dealing with that. I get furious when I see the comment section and feel like writing them, "Hey, what's going on here, why are you allowing this comment? Why don't you block this comment or delete it or whatever!?" (Franziska, 25 years)

In addition to construing personal responsibility along the boundaries of one's personal public sphere, a lack of urgency to undertake impression management in settings where their actions were not observable by imagined audiences often facilitated inaction. Friedrich explained it this way:

> I scroll through my news feed, see [something uncivil], don't like what I see, but nobody sees that I was there. I don't feel like society expects me to step in there. But, for example, on WhatsApp, people see that I could have reacted to it, so I have to intervene there. Otherwise, they might think I agree with an opinion because of my passive behavior. Or they might judge me: "Why didn't you react to that if you disagree?"

This is not to say that participants disapproved of bystander intervention. Rather, most believed it was legitimate for bystanders to disengage, displacing responsibility onto the user, page, or a group of users perceived as responsible for a particular fragment of the public sphere.

Discussion

The present study investigated how Facebook users construe personal responsibility to intervene against incivility. In a field dominated by quantitative surveys and experimental research, this study offers a sociological and constructivist take on user intervention

in that it foregrounds the *social* in social media. Specifically, I explored how social, spatial, and situational aspects of everyday social media matter for users' understandings of personal responsibility to intervene against incivility.

The study reveals that "civic territories along which citizens understand and practice their civic duties" (Papacharissi, 2010: 16) differ from scholarly modes of imagining the public sphere and formulating expectations of bystander intervention against incivility. As multiple personal public spheres intersect in users' news feeds, rather than being responsible for intervention everywhere and at all times, those perceived as sovereigns in a delimited communicative space—their personal public sphere—are most strongly expected to intervene. Users are considered personally accountable for managing their personal public sphere to the best of their ability, enforcing norms that they consider worthwhile. This includes not only exercising invisible sanctions, such as unfriending (John and Agbarya, 2021; John and Gal, 2018), but also publicly silencing uncivil users (Tsfati and Dvir-Gvirsman, 2018). The pressure to react to incivility in one's personal public sphere—where their intervention (or lack thereof) is publicly visible, and supervision by social ties is relatively high—seems to thwart the bystander effect by strengthening individuals' public self-awareness (Van Bommel et al., 2012).

Social relatedness with uncivil users extends the idea of responsibility from discursive to relational concerns (Gagrčin et al., 2022; Gagrčin and Porten-Cheé, 2022), grounding the sense of personal responsibility to intervene in the relationship one has with the person rather than in the horizontal nature of civic relations (Dishon and Ben-Porath, 2018). Social relatedness to uncivil users induces a hierarchization of responsibility to enforce norms and shapes the quality of intervention (from sanctioning to reforming). Aiming to sustain the relationship by "clearing the air" (McLaughlin and Vitak, 2012: 311), users seem comparably more likely to engage in some sort of confrontation with close social ties. The relevance of social relatedness is evident also in the perceived appropriateness of intervention—the final step preceding the act of intervention in the bystander intervention model. A close social connection with an uncivil user does not relieve users of the responsibility to intervene publicly but prompts them to insulate the reforming part of intervention from unwanted audiences by moving it to the virtual backstage.

The present study challenges the scholarly fixation on news media comment sections as central spaces for intervention on social media by highlighting personal public spheres as spaces of meaningful social influence. Thus, instead of treating users as social aggregates, it becomes apparent that in the context of mediated social life, user intervention is not an isolated act of flagging or counter-speaking but a highly contextual matter with real consequences in the life worlds of users (Morey et al., 2012; Neubaum et al., 2021a, 2021b). In this light, the study shifts the focus from bystanders as intervening actors to proprietors of personal public spheres.

Thereupon, I suggest an alternative frame of user responsibility in social media environments. Moe (2020) argues that in the light of contemporary information abundance, digital citizenship "cannot be assessed based on individual citizens in isolation, but should be considered as distributed, and embodied in citizens' social networks, with a division of labor" (p. 1). Given the amount of disruptive content on social platforms, I show that users rely on heuristics such as a delimited space of responsibility or the

involvement of meaningful social ties to determine when and how they are expected to intervene. Thus, building on Moe's concept of "distributed readiness citizenship," individual responsibility for enforcing norms by intervening against incivility can be understood as *distributed among personal public spheres* (cf. Draper, 2019). When responsibility is clearly attributed to the proprietor of a personal public sphere, intervention becomes immediately important to the person in question because instances of incivility create impression and relational management urgencies. Reframing responsibility as distributed in this way takes into account that users negotiate their role in the public discourse "via the nexus of a private sphere" (Papacharissi, 2010: 24), where social and civic responsibilities frequently overlap and are difficult to distinguish (Sinclair, 2012). In this sense, it enables us to consider the relevance of citizens' social ties for enforcing norms in the public discourse online (Moisuc and Brauer, 2019; Sinclair, 2012)—an aspect thus far under-researched in the field of user intervention but likely to gain prominence as informal political talk online increasingly moves into chat groups.

On a critical note, this study is limited to only one platform, and how users imagine and draw boundaries of their personal public sphere is likely to differ between platforms, contingent upon perceived norms and affordances (Literat and Kligler-Vilenchik, 2021). For example, platform-specific constellations that permit the mutual observability of actors and audiences, such as chat groups, may be more likely to produce a sense of personal responsibility for bystander intervention (e.g. Kligler-Vilenchik, 2022). Moreover, the adopted methodological approach to eliciting norms and expectations was admittedly likely to produce social desirability. Nonetheless, it is remarkable that respondents reasoned against a personal responsibility to intervene, adding nuance to the abstract idea of responsibility, which was arguably the study's intention. Since the type of victim matters for perceptions of personal responsibility (Naab et al., 2018), it is also relevant to highlight that while I studied incivility directed toward social groups, the characters in the vignettes were explicitly not members of the targeted groups—an aspect that several participants mentioned as a condition for placing the responsibility to act on the proprietor of the personal public sphere.

Although the results support research conducted elsewhere (e.g. Schwarz and Shani, 2016; Tsfati and Dvir-Gvirsman, 2018), the specificity of the context should be noted, particularly since the value of civil courage rates high in Germany. Future research could address these questions from a comparative perspective (e.g. platforms, countries), empirically test the propositions made in this study in an experimental design (e.g. using the bystander intervention framework and varying the degree of social proximity to the uncivil users), and inquire how different groups (e.g. minorities typically targeted by incivility, illiberal individuals, other age groups) conceive of user responsibility to fight incivility in diverse situational settings. Future research would benefit from media sociological perspectives that treat interaction in social media environments as socially embedded and contextual.

Conclusion

Amid growing concerns about incivility on social media and deficient platform moderation practices, democratic discourse on social media platforms depends on ordinary users' sense of personal responsibility to (re)assert norms. This study shows that the

boundaries of one's personal public sphere and social relatedness to uncivil users serve as heuristics for thinking about their personal responsibility to intervene against incivility. The presence of relevant social ties and the desire to maintain face compel users to engage in intervening behavior. Counter to the popular focus on news comments as relevant sites for user intervention, users perceive their personal public spheres as comparably more important, efficacious, and appropriate sites for norm enforcement and peer influence. In the absence of personal responsibility for news media comment sections, the results underscore the need for organized comment moderation on news media outlets' pages. If we are to foster civic intervention against incivility, we ought to employ more person-centric (in addition to discourse-centric) and socially embedded approaches to users' roles in online public discourse. Reimagining user responsibility as distributed among personal public spheres is one way of delimiting the space of individual responsibility, making user intervention not only immediately important but also practically feasible in the context of everyday social media use.

Acknowledgements

A big thank you to my student assistants, Miriam Milzner, Paula Starke, Sofie Jokerst, and Christina Hecht who supported me in recruiting participants and transcribing the interviews and to my colleagues Laura Leißner and Marlene Kunst who provided useful information in the initial construction of the vignettes. I am also very grateful to Neta Kligler-Vilenchik, Juliane Lischka, Martin Emmer, and Nadja Schaetz who provided valuable feedback on earlier versions of the manuscript. Finally, I am thankful to the anonymous reviewers for their suggestions for improving this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the German Federal Ministry of Education and Research, grant number 16DII114.

ORCID iD

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

Supplemental material

Supplemental material for this article is available online.

Note

1. Nevertheless, I may note that what participants bemoaned as "intervening for the record" in fact contributes to a more civil and deliberative discourse because it signals descriptive norms (Friess et al., 2020).

References

Andersen K, Ohme J, Bjarnøe C, et al. (2020) *Generational Gaps in Political Media Use and Civic Engagement*. Abingdon: Routledge.

Anderson AA, Brossard D, Scheufele DA, et al. (2014) The "nasty effect": online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication* 19(3): 373–387.

Barnidge M and Rojas H (2014) Hostile media perceptions, presumed media influence, and political talk: expanding the corrective action hypothesis. *International Journal of Public Opinion Research* 26(2): 135–156.

Baym NK and boyd d (2012) Socially mediated publicness: an introduction. *Journal of Broadcasting & Electronic Media* 56(3): 320–329.

Bormann M, Tranow U, Vowe G, et al. (2021) Incivility as a violation of communication norms—a typology based on normative expectations toward political communication. *Communication Theory* 00: 1–31.

Buerger C (2021) #iamhere: collective counterspeech and the quest to improve online discourse. *Social Media + Society* 7(4): 1–17.

Cialdini RB, Kallgren CA and Reno RR (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.

Dahlgren P (2006) Doing citizenship: the cultural origins of civic agency in the public sphere. *European Journal of Cultural Studies* 9(3): 267–286.

Dishon G and Ben-Porath S (2018) Don't @ me: rethinking digital civility online and in school. *Learning, Media and Technology* 43(4): 434–450.

Donath J (2007) Signals in social supernets. *Journal of Computer-Mediated Communication* 13(1): 231–251.

Draper NA (2019) Distributed intervention: networked content moderation in anonymous mobile spaces. *Feminist Media Studies* 19(5): 667–683.

Emmer M, Leißner L, Strippel C, et al. (2021) *Weizenbaum Report 2021: Politische Partizipation in Deutschland* (Political participation in Germany, Report no. 2). Berlin: Weizenbaum Institute for the Networked Society.

Eriksson K, Andersson PA and Strimling P (2017) When is it appropriate to reprimand a norm violation? The roles of anger, behavioral consequences, violation severity, and social distance. *Judgment and Decision Making* 12(4): 396–407.

Festinger L (1957) *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson and Company.

Freelon D (2015) Discourse architecture, ideology, and democratic norms in online political discussion. *New Media & Society* 17(5): 772–791.

Friess D, Ziegele M and Heinbach D (2020) 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.

Gagrčin E and Porten-Cheé P (2022) Individual and collective social effort: ideals and practices of informed citizenship in different information environments. *International Journal of Communication* 16: 1–20.

Gagrčin E, Porten-Cheé P, Leißner L, et al. (2022) What makes a good citizen online? The emergence of discursive citizenship norms in social media environments. *Social Media + Society* 8(1): 1–11.

Giansiracusa N (2021) Facebook uses deceptive math to hide its hate speech problem. *Wired*, 15 October. Available at: https://www.wired.com/story/facebooks-deceptive-math-when-it-comes-to-hate-speech/

Gillespie T (2020) Content moderation, AI, and the question of scale. *Big Data and Society* 7(2): 1–5.

Goffman E (1967) *Interaction Ritual: Essays in Face-to-Face Behavior*. New York: Pantheon.

Gray D, Royall B and Malson H (2017) Hypothetically speaking: using vignettes as a standalone qualitative method. In: Braun V, Clarke V and Gray D (eds) *Collecting Qualitative Data: A Practical Guide to Textual, Media and Virtual Techniques*. Cambridge: Cambridge University Press, pp. 45–70.

Hall JA (2012) Friendship standards: the dimensions of ideal expectations. *Journal of Social and Personal Relationships* 29(7): 884–907.

Heger K, Leißner L, Emmer M, et al. (2022) *Weizenbaum Report 2022: Politische Partizipation in Deutschland* [Political participation in Germany, Report no. 3]. Berlin: Weizenbaum Institute for the Networked Society.

Heldt A (2019) Reading between the lines and the numbers: an analysis of the first NetzDG reports. *Internet Policy Review* 8(2). DOI: 10.14763/2019.2.1398.

Hofmann W, Brandt MJ, Wisneski DC, et al. (2018) Moral punishment in everyday life. *Personality and Social Psychology Bulletin* 44(12): 1697–1711.

Hogan B (2010) The presentation of self in the age of social media: distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society* 30(6): 377–386.

John NA and Agbarya A (2021) Punching up or turning away? Palestinians unfriending Jewish Israelis on Facebook. *New Media & Society* 23(5): 1063–1079.

John NA and Gal N (2018) "He's got his own sea": political Facebook unfriending in the personal public sphere. *International Journal of Communication* 12: 2971–2988.

Kalch A and Naab TK (2017) Replying, disliking, flagging: how users engage with uncivil and impolite comments on news sites. *Studies in Communication and Media* 6(4): 395–419.

Kim Y (2021) Understanding the bystander audience in online incivility encounters: conceptual issues and future research questions. In: *Proceedings of the 54th Hawaii international conference on system sciences*, 5–8 January, pp. 2934–2943. Available at: http://hdl.handle.net/10125/70971

Kligler-Vilenchik N (2021) Friendship and politics don't mix? The role of sociability for online political talk. *Information, Communication & Society* 24(1): 118–133.

Kligler-Vilenchik N (2022) Collective social correction: addressing misinformation through group practices of information verification on WhatsApp. *Digital Journalism* 10: 300–318.

Kunst M, Porten-Cheé P, Emmer M, et al. (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.

Latané B and Darley JM (1970) *The Unresponsive Bystander: Why Doesn't He Help?*. Hoboken, NJ: Prentice Hall.

Lieberman D and Linke L (2007) The effect of social category on third party punishment. *Evolutionary Psychology* 5(2): 289–305.

Literat I and 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): 1–14.

Litt E (2012) Knock, knock: who's there? The imagined audience. *Journal of Broadcasting and Electronic Media* 56(3): 330–345.

McLaughlin C and Vitak J (2012) Norm evolution and violation on Facebook. *New Media & Society* 14(2): 299–315.

Marwick AE and boyd d (2011) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13(1): 114–133.

Masullo GM and Kim J (2021) Exploring "angry" and "like" reactions on uncivil Facebook comments that correct misinformation in the news. *Digital Journalism* 9(8): 1103–1122.

Masullo GM, Muddiman A, Wilner T, et al. (2019) We should not get rid of incivility online. *Social Media + Society* 5(3): 1–5.

Meta (2021) Hate speech prevalence has dropped by almost 50% on Facebook. Available at: https://about.fb.com/news/2021/10/hate-speech-prevalence-dropped-facebook/ (accessed 13 June 2022).

Meta (2022) *Facebook Community Standards: Hate Speech*. Facebook Transparency Center. Available at: https://transparency.fb.com/de-de/policies/community-standards/hate-speech/ (accessed 13 June 2022).

Moe H (2020) Distributed readiness citizenship: a realistic, normative concept for citizens' public connection. *Communication Theory* 30(2): 205–225.

Moisuc A and Brauer M (2019) Social norms are enforced by friends: the effect of relationship closeness on bystanders' tendency to confront perpetrators of uncivil, immoral, and discriminatory behaviors. *European Journal of Social Psychology* 49(4): 824–830.

Mor Y, Kligler-Vilenchik N and Maoz I (2015) Political expression on Facebook in a context of conflict: dilemmas and coping strategies of Jewish-Israeli youth. *Social Media + Society* 1(2): 1–10.

Morey AC, Eveland WP and Hutchens MJ (2012) The "who" matters: types of interpersonal relationships and avoidance of political disagreement. *Political Communication* 29(1): 86–103.

Naab TK, Kalch A and Meitz TGK (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.

Neubaum G, Cargnino M and Maleszka J (2021a) How Facebook users experience political disagreements and make decisions about the political homogenization of their online network. *International Journal of Communication* 15: 187–206.

Neubaum G, Cargnino M, Winter S, et al. (2021b) "You're still worth it": the moral and relational context of politically motivated unfriending decisions in online networks. *PLoS ONE* 16(1): e0243049.

O'Dell L, Crafter S, de Abreu G, et al. (2012) The problem of interpretation in vignette methodology in research with young people. *Qualitative Research* 12(6): 702–714.

Papacharissi Z (2010) *A Private Sphere: Democracy in a Digital Age*. Malden, MA: Polity.

Passy F and Monsch GA (2020) *Contentious Minds: How Talk and Ties Sustain Activism*. New York: Oxford University Press.

Pearce KE, Vitak J and Barta K (2018) Socially mediated visibility: friendship and dissent in authoritarian Azerbaijan. *International Journal of Communication* 12: 1310–1331.

Porten-Cheé P, Kunst M and Emmer M (2020) Online civic intervention: a new form of political participation under conditions of a disruptive online discourse. *International Journal of Communication* 14: 21–21.

Reichelmann A, Hawdon J, Costello M, et al. (2021) Hate knows no boundaries: online hate in six nations. *Deviant Behavior* 42(9): 1100–1111.

Riedl MJ, Naab TK, Masullo GM, et al. (2021) Who is responsible for interventions against problematic comments? Comparing user attitudes in Germany and the United States. *Policy and Internet* 13(3): 433–451.

Rossini P (2022) Beyond incivility: understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research* 49(3): 399–425.

Saldaña J (2016) *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: SAGE.

Saldaña J and Omasta M (2018) *Qualitative Research: Analyzing Life*. Los Angeles, CA: SAGE.

Schaetz N, Leißner L, Porten-Cheé P, et al. (2020) *Weizenbaum Report 2020: Politische Partizipation in Deutschland* [Political participation in Germany, Report no. 1]. Berlin: Weizenbaum Institute for the Networked Society.

Schmidt JH (2014) Twitter and the rise of personal publics. In: Weller K, Bruns A, Burgess J, et al. (eds) *Twitter and Society*. New York: Peter Lang, pp. 3–14.

Schwarz O and Shani G (2016) Culture in mediated interaction: political defriending on Facebook and the limits of networked individualism. *American Journal of Cultural Sociology* 4(3): 385–421.

Siapera E and Viejo-Otero P (2021) Governing hate: Facebook and digital racism. *Television and New Media* 22(2): 112–130.
Sinclair B (2012) *The Social Citizen*. Chicago, IL: University of Chicago Press.
Strimling P and Eriksson K (2014) Regulating the regulation. In: Van Lange P, Rockenbach B and Yamagishi T (eds) *Reward and Punishment in Social Dilemmas*. New York: Oxford University Press, pp. 52–69.
Stroud NJ, Scacco JM, Muddiman A, et al. (2015) Changing deliberative norms on news organizations' Facebook sites. *Journal of Computer-Mediated Communication* 20(2): 188–203.
Thorson K (2014) Facing an uncertain reception: young citizens and political interaction on Facebook. *Information Communication and Society* 17(2): 203–216.
Thorson K and Wells C (2016) Curated flows: a framework for mapping media exposure in the digital age. *Communication Theory* 26(3): 309–328.
Timberg C (2021) New whistleblower claims Facebook allowed hate, illegal activity to go unchecked. *The Washington Post*, 22 October. Available at: https://www.washingtonpost.com/technology/2021/10/22/facebook-new-whistleblower-complaint/
Tsfati Y and Dvir-Gvirsman S (2018) Silencing fellow citizens: conceptualization, measurement, and validation of a scale for measuring the belief in the importance of actively silencing others. *International Journal of Public Opinion Research* 30(3): 391–419.
Van Bommel M, Van Prooijen JW, Elffers H, et al. (2012) Be aware to care: public self-awareness leads to a reversal of the bystander effect. *Journal of Experimental Social Psychology* 48(4): 926–930.
Vraga EK, Thorson K, Kligler-Vilenchik N, et al. (2015) How individual sensitivities to disagreement shape youth political expression on Facebook. *Computers in Human Behavior* 45: 281–289.
Wang S and Kim KJ (2020) Restrictive and corrective responses to uncivil user comments on news websites: the influence of presumed influence. *Journal of Broadcasting and Electronic Media* 64(2): 173–192.
Watson BR, Peng Z and Lewis SC (2019) Who will intervene to save news comments? Deviance and social control in communities of news commenters. *New Media & Society* 21(8): 1840–1858.
Wells C and Thorson K (2017) Combining big data and survey techniques to model effects of political content flows in Facebook. *Social Science Computer Review* 35(1): 33–52.
Wilhelm C, Joeckel S and Ziegler I (2020) Reporting hate comments: investigating the effects of deviance characteristics, neutralization strategies, and users' moral orientation. *Communication Research* 47(6): 921–944.
Wilks T (2004) The use of vignettes in qualitative research into social work values. *Qualitative Social Work* 3(1): 78–87.
Wintterlin F, Frischlich L, Boberg S, et al. (2021) Corrective actions in the information disorder: the role of presumed media influence and hostile media perceptions in the countering of distorted user-generated content. *Political Communication* 38(6): 773–791.
Ziegele M, Naab TK and Jost P (2020) Lonely together? Identifying the determinants of collective corrective action against uncivil comments. *New Media & Society* 22(5): 731–751.

Author biography

Emilija Gagrčin (MA, Freie Universität Berlin) is a doctoral student at the Freie Universität Berlin and a research associate at the Weizenbaum Institute for the Networked Society. Her research interests include social and normative aspects of political communication in digital environments.