Article

Social Media Users' Motives for (Not) Engaging With Hate Speech: An Explorative Investigation

Ursula Kristin Schmid (iD), Anna Sophie Kümpel (iD),
and Diana Rieger (iD)

Abstract

Despite extensive research on what causes social media users to recognize hate speech and what motivates their reactions to it, little is known about a crucial intermediate step that leads to users' engagement or non-engagement with hate speech online. In our study, drawing on the uses and gratifications approach, we theoretically derived motives representing affective and entertainment, personal identity and social-integrative as well as cognitive dimensions for social media users to engage or not engage with hate speech. To empirically investigate those motives, we conducted a quota-based online survey of adult social media users in Germany ($N = 4,020$) and subjected the responses to exploratory factor analysis. We found that a range of personal and social gratifications, far beyond simple expressions of approval or disapproval, encourage social media users to engage with hate speech online, whereas intentions to protect oneself and others from potential harm discourage such behavior.

Keywords

hate speech, engagement, uses and gratification, social media, exploratory factor analysis

Online hate speech is hardly rare but instead becoming increasingly the "norm" for internet users, particularly younger ones (Haslop et al., 2021). Such online hate speech, regarded as an extreme form of online incivility, manifests in public expressions of hatred or degrading attitudes toward social groups. When people come across hate speech online, they can respond in one of two ways: one, engage with it by carefully examining its problematic content and contemplating its underlying message and potential consequences or, two, simply ignore it and not spend any more time on it. In a recent survey differentiating those response options, 76% of internet users in Germany reported having encountered hate speech online, although only 40% of them reported engaging with it in some way (Landesanstalt für Medien NRW, 2023). However, those numbers do not explain the actual reasons for engaging or not engaging with hate speech online.

Social media users' engagement with hate speech can have both positive and negative outcomes not only for themselves but also for society as a whole. For example, exposure to hate speech has been shown to adversely impact individuals' emotions (Gervais, 2015) and mental health (Cover, 2023; Gelber & McNamara, 2016). However, it can also have positive effects by, for instance, entertaining or

generating enthusiasm (Kosmidis & Theocharis, 2020; Sydnor, 2018). At the societal level, engaging with hate speech and reflecting problematic content can also be both advantageous and disadvantageous. For example, if users decide to counter hate speech by writing an intervening comment, then it could take the discussion in both a favorable and detrimental direction (Schäfer et al., 2023).

In our research, we conceived those possible outcomes as (personal or societal) gratifications or inconveniences for social media users that guide their decisions of whether or not to engage with hate speech online. Adopting a uses and gratifications perspective, we delineated various motivations for and against engaging with hate speech, all of which we exploratively derived from recent research on factors influencing perceptions of hate speech, bystander intervention, and the perpetration of hate speech. We also drew on literature regarding the umbrella concept of incivility as

LMU Munich, Germany

Corresponding Author:
Ursula Kristin Schmid, Department of Media and Communication, LMU Munich, Oettingenstr. 67, Munich 80538, Germany.
Email: ursula.schmid@ifkw.lmu.de
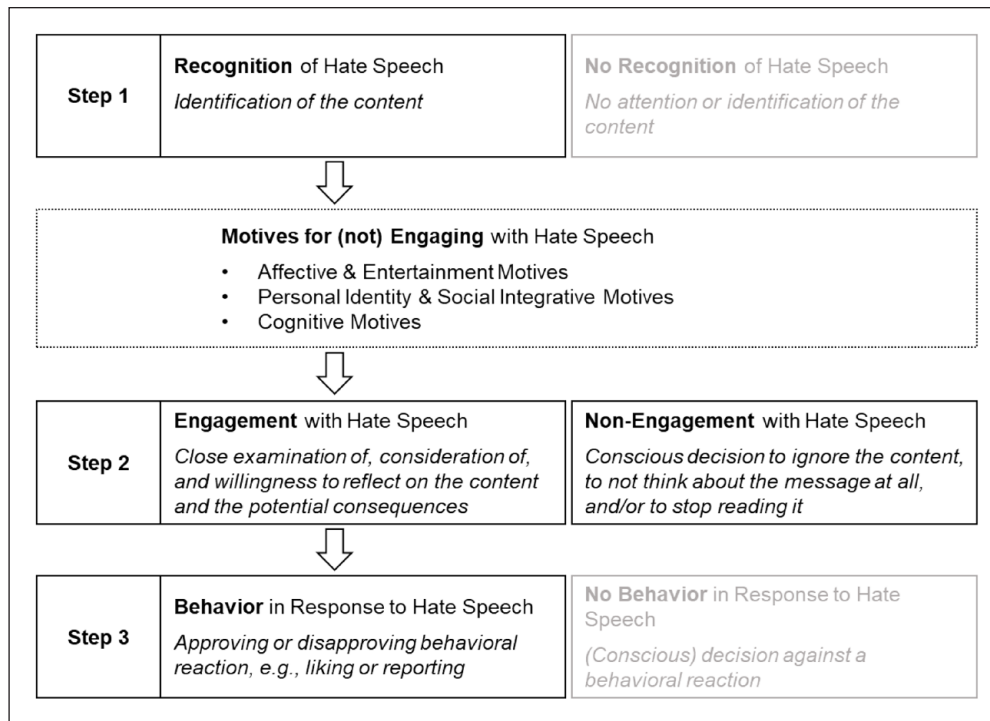
Figure 1. Multistep process of social media users' reception of hate speech.

norm-violating communication to map a broad spectrum of motives. To the same purpose, we also reviewed studies on users' motives for engaging with online content in general. As a result, we gain insights into not only motives for consciously engaging with hate speech but also motives for consciously not engaging, a frequently neglected aspect that is nevertheless crucial for effectively addressing hate speech. To empirically examine users' motives for both responses, we conducted a large-scale, quota-based online survey representative of adult social media users in Germany ($N=4,020$).

Hate Speech on Social Media Platforms and Users' Reception

In line with previous research (Hawdon et al., 2017; Rieger et al., 2020; Schmid, Kümpel, & Rieger, 2024), we define *hate speech* as public expressions of hate or degrading attitudes toward a social group or collective. People targeted by hate speech are devalued based on group-defining characteristics such as ethnicity, religion, and sexual orientation. As such, hate speech is considered an extreme and group-related form of *incivility*, defined as comprising "features of discussion that convey an unnecessarily disrespectful tone" (Coe et al., 2014, p. 660) and serving as an umbrella term for communication that violates one or multiple norms (Bormann et al., 2021). Social media platforms usually contain a mixture of norm-violating speech, ranging from blatant uncivil language that violates mutual respect norms on a linguistic level to (more or less overt) forms of devaluation that violate

norms of democratic tolerance on a substantive level (Rossini, 2022). To provide a comprehensive understanding of the motives for and against engaging with hate speech as a form of group-related antinormative online communication, we refer to both literature that deals explicitly with hate speech and with the broader concept of incivility in the sections that follow.

Although hate speech is not a phenomenon unique to social media, the features of social media platforms and their users make its dissemination and propagation especially straightforward on social media (Matamoros-Fernández, 2017). As a result, even indifferent users might encounter hate speech or hateful ideology simply while scrolling through their social media feeds (Matamoros-Fernández, 2017, 2018).

When internet users confront hate speech, their response involves a multistep process that requires the clear differentiation of the steps entailed. On that count, Sydnor (2018) has separated the "identification of uncivil cues and assessment of the message as civil or uncivil, emotional engagement with the message, and subsequent feelings that the message is interesting or entertaining" (p. 100) as consecutive but distinct steps in perceiving incivility. Similarly, we differentiate (1) the recognition of hate speech, (2) engagement with said speech, and (3) subsequent behavior (see Figure 1). Although influencing factors of and motives for the first and third steps (i.e., recognition and behavior) have been systematically examined, research has largely overlooked the step in-between—that is, users' decision about whether to engage with content recognized as hate speech.

Research on the recognition of social media content reveals that as individuals scroll through social media platforms, their awareness varies, thereby meaning that not all content, hate speech included, is consistently recognized (Schmid, Kümpel, & Rieger, 2024). Studies have demonstrated that the form in which hate speech appears (e.g., textual, visual, and audiovisual) and its distinctive characteristics (e.g., severity of hate speech) influence its recognition, with visual and extreme speech being more likely to capture attention (Schmid, Kümpel, & Rieger, 2024; Sydnor, 2018). Regarding users' behavior in response to hate speech, including engaging in counter speech or reporting it to platform's operators, various factors have been identified as influential, including personal (political) characteristics, norms, and attitudes, as well as the severity of the hate speech and the identity of the targeted group (Gagrčin & Milzner, 2023; Leonhard et al., 2018; Schmid, Obermaier, & Rieger, 2024). For instance, social media users are generally more motivated to intervene against online hate speech if they perceive the content as being particularly threatening, feel personally responsible for acting against it (Leonhard et al., 2018), or perceive themselves as being able to achieve something as a result of their (political) actions (Schmid, Obermaier, & Rieger, 2024). However, that decision has to result from a somewhat in-depth engagement with the content, which does not always take place, even if the hate speech is indeed recognized in the first step (Schmid, Kümpel, & Rieger, 2024). Thus, in our study, we sought to systematically examine the motives that influence social media users' decisions during the overlooked second step—that is, to engage or not engage with hate speech online.

In contrast to *recognition of hate speech*, meaning users' mere identification of the content upon first contact, we define *engagement with hate speech* as the close examination of, consideration of, and willingness to reflect on hateful messages and their potential consequences.[1] By extension, *non-engagement with hate speech* implies choosing to consciously ignore such speech, to not think about the message at all, and/or to stop reading it. The distinction between engagement and non-engagement is not strictly binary but spans a spectrum ranging from total avoidance, in which users stop reading entirely, through moderate engagement, which involves a degree of cognitive interaction with the content, to deep engagement and reflection. In our study, we assumed that different motives may be crucial for users' (extent of) engagement, which, however, have not been explored empirically yet. Our research question was, therefore, the following:

RQ: What motives do social media users have for (a) engaging and (b) not engaging with hate speech?

Possible Motives for (Not) Engaging With Hate Speech

The use of different media and media content is often associated with users' expected gratifications and empirically examined by building on the uses and gratifications approach (Katz et al., 1973). According to the approach, media users are aware of their own needs and actively select media (content) in ways that satisfy those needs. As a result of their selection of content, users may in fact be rewarded with those gratifications. In their early study, Katz et al. (1973) explored four basic dimensions of needs in terms of media consumption: a cognitive dimension, including information seeking; an affective dimension, including emotional experiences and entertainment; a social-integrative dimension, including social exchange; and a personal identity dimension, including self-affirmation. Similarly, McQuail (1987) classified recipients' motives for using media into four major categories: entertainment, integration and social interaction, personal identity, and information. With reference to those overarching dimensions, researchers have extensively examined the motives for engaging with various types of media (content) in recent years, also by focusing on what motivates users to engage with online content such as YouTube videos (Khan, 2017) or other users' comments (Diakopoulos & Naaman, 2011; Springer et al., 2015). In another classification, Khan (2017) has categorized such user engagement into two distinct forms—participation, exemplified by activities such as writing comments, and consumption, highlighted by activities such as reading comments—both associated with different types of gratifications.

In this article, we focus on the consumption form of engagement with the aim to understand users' motives for engaging with a particular type of content: online hate speech. We assume that the choice to engage with hate speech is a deliberate and conscious one and contrasts the initial recognition of such content. For that reason, we refer to the uses and gratifications approach to describe the gratifications that social media users seek from engaging with hate speech online. Our approach aligns with prior research that has investigated users' motives for engaging with online comments (Diakopoulos & Naaman, 2011; Khan, 2017; Springer et al., 2015). Moreover, we assume that even non-engagement is an active decision and thus subject to similar motives that may nevertheless be driven more by a desire to avoid discomfort than to be gratified (see also Springer et al., 2015; Stroud et al., 2016).

Thus, in the following sections, we outline possible motives for both consciously engaging and not engaging with hate speech upon recognizing it, categorized into (1) affective and entertainment motives, (2) personal identity and social-integrative motives, and (3) cognitive motives (see also Diakopoulos & Naaman, 2011; Springer et al., 2015). By

comparison, Niehoff and Oosterwijk (2020) have suggested examining curiosity about negative content—which could include hate speech—in terms of emotional, social, and informational motives. We also apply empirical insights from research on reading users' comments in general, the perception of negative content and hate speech, interventions against hate speech, and the perpetration of hate speech as well as discuss how influencing factors in those areas might also be crucial for users' motivation to (not) engage with such speech. In doing so, we acknowledge that several motives can exist simultaneously and/or vary among users depending on the situation, just as users' behavior in response to hate speech can be guided by several personal norms at the same time (Gagrčin & Milzner, 2023).

Affective and Entertainment Motives

Users react to hate speech in different ways and with a range of emotions and degrees of intensity—influenced, among others, by factors such as the type and severity of the content, as well as personal attitudes (Schmid, Kümpel, & Rieger, 2024). For most users, being confronted with hate speech threatens personal well-being and causes negative emotions such as anger and anxiety (Landesanstalt für Medien NRW, 2023; Sydnor, 2018), as well as shock and concern (Schmid, Kümpel, & Rieger, 2024). Social media users who find themselves within the group targeted by hate speech are particularly prone to negative psychological impacts (Cover, 2022; Gelber & McNamara, 2016), for aggression is perceived to be more offensive when directed at one's own group, especially when similar experiences have occurred in the past (Williams et al., 2016). In particular, exposure to uncivil political talk that the user disagrees with has been found to induce anger, aversion, and diminished satisfaction with participating in the discourse (Gervais, 2015).

Such negative emotional affect can lead to divergent behaviors. On the one hand, people may consciously choose to disengage from the content and/or even avoid political talk in general (Barnidge et al., 2019) out of self-protection and to avoid emotional distress (Gagrčin & Milzner, 2023; Schmid, Kümpel, & Rieger, 2024). In that regard, a survey of U.S. citizens has revealed that people avoid reading users' comments mostly because the comments are argumentative (41% agreement) and mean-spirited or uncivil (40% agreement; Stroud et al., 2016). Thus, consciously ignoring the content to avoid negative personal outcomes might be reasonable for many social media users. On the other hand, negative emotional affect could be associated with engaging even more deeply—for example, if users feel duty-bound to intervene (Leonhard et al., 2018) and mitigate the threat, thereby reducing negative emotions.

By contrast, for some users, negative emotions such as fear or sadness can elicit sensations and arousal that they value, thereby leading them to actively seek out such pleasurable "hedonic reversals" (Rozin et al., 2013) within aversive content, which may include conflicts and/or violence. Regarding political disputes on TV, evidence suggests that such content's perceived entertainment value is greater for uncivil disputes than for civil ones (Mutz & Reeves, 2005). Uncivil political exchanges on Twitter have also been found to enthuse and/or entertain some recipients (Kosmidis & Theocharis, 2020; Sydnor, 2018). Moreover, similar to Erjavec and Kovačič's (2012) group of "players" in their categorization of hate speech perpetrators who spread hate speech primarily for the thrill and fun of it, some users also engage with hate speech primarily for the purposes of hedonic entertainment ("fun") and even intentionally seek out hateful, sensationalist content (Bedrosova et al., 2023). Likewise, in evidence of seeking and/or experiencing enjoyment and fun from reading users' comments in general (Diakopoulos & Naaman, 2011; Stroud et al., 2016), strong motives for entertainment have been shown to increase how often one reads comments (Springer et al., 2015). Even those studies, which have explored engagement with users' comments without focusing on hateful content, have demonstrated that the motive for entertainment is usually tied to conflict, such as enjoying debates with others (Diakopoulos & Naaman, 2011) or finding it "entertaining to see others fight" (Springer et al., 2015, p. 807).

In a similar vein, consuming incivility, hate, and other "dark content," so to speak, can induce eudaimonic entertainment experiences (e.g., Bartsch et al., 2016). Eudaimonic entertainment, usually linked to morally and meaningful charged content, stirs emotional responses that include sentiments of inspiration, gratitude, and elevation (e.g., Oliver & Raney, 2011). For instance, extremist content, including extremist Islamic posts on Instagram, often entail eudaimonic and inspirational elements such as depictions of nature or moral virtue (Frischlich, 2021). Consuming such media content, even if violent, can induce reflections on meaning, morality, and authenticity (Bartsch et al., 2016) and even increase prosocial intentions (Grizzard et al., 2017). Along similar lines, encountering hate speech and engaging with it may trigger a sense of eudaimonic entertainment, just as expressing hate toward other groups has been found to have the potential to reduce threats and increase meaning in life (Elnakouri et al., 2022).

Personal Identity and Social-Integrative Motives

Feeling entertained by hate speech is far more likely if users agree with the opinions expressed and if the hate speech thus connects to their social identity. Being exposed to like-minded incivility has been shown to increase the use of own uncivil behavior and utterances (Gervais, 2015), which could be understood as an outcome of more in-depth engagement with such content. In those cases, dealing with hate speech in greater depth may simply afford a means to confirm one's opinion and find like-minded peers. Just as mutual social support, attention, and approval from one's in-group serve as

motivators for posting hate speech (Walther, 2022), similar intergroup dynamics may prompt engagement with hate speech on social media posted by like-minded individuals. Again, evidence already suggests a positive association between reading users' comments in general and personal identity motives that include "validating or comparing one's opinion against that of the community" (Diakopoulos & Naaman, 2011, p. 138) or checking "to see how your opinion of the story or topic compares to others' view" (Stroud et al., 2016, p. 10).

Engaging with hate speech may also result from being personally involved because the speech represents a threat to one's ingroup and identity or threatens close others. Similar arguments have previously been made regarding intervention behavior in response to hate speech that is considered to be collective action (Ziegele et al., 2020). From that perspective, collective benefits to the community that contribute to a climate of respectful, deliberative conversation can drive people to intervene (Jost et al., 2020). That dynamic may be particularly true for people with high levels of empathy, an emotion shown to be positively associated with engaging in helping behavior after witnessing cyberbullying (Hu et al., 2023). Especially when identifying with people under attack, engaging with hate speech can lead to empathy and sharing feelings with those people by taking their perspectives, which can consequently foster a sense of social belonging (see Niehoff & Oosterwijk, 2020). In general, social media users are also more likely to intervene if they know the author(s) or the target(s) personally (Gagrčin, 2022). Prior to that, being personally involved to some degree might prompt higher levels of engagement with hate speech, whereas the lack of social ties might lower perceived relevance and subsequent engagement.

The desire for social interaction and participation in online discussions in general may be another motivational aspect, for maintaining or establishing social conversations and relationships is a common motive for using media as identified in research following the uses and gratifications approach (Katz et al., 1973; McQuail, 1987). Regarding online content, research has suggested that comments from other users can serve as an indicator of public opinion (Zerback & Fawzi, 2017) and the pulse of the community. Comments are also a valuable resource for follow-up communication, even extended into offline contexts (Springer, 2014), which is a major reason why people engage with them (Diakopoulos & Naaman, 2011; Springer et al., 2015; Stroud et al., 2016). That motive might be particularly relevant in regard to emotionally charged content such as hate speech, which attracts considerable attention and provides material for discussion. In that light, following emotive—in our case, hateful—online discussions may simply mean being involved with the liveliest topics online, staying informed, and being able to converse with others.

However, the awareness that hate speech receives significant attention and can spread hostility due to its emotionally charged nature may also prompt individuals to deliberately ignore the content or consciously distance themselves from it. Along with avoiding emotional distress on the individual level, refraining from engaging with hate speech might serve to prevent harm on a societal level (e.g., negative polarizing effects) that can arise from giving hate speech attention (Schäfer et al., 2023).

Cognitive Motives

Closely connected to the desire of being able to participate in (political) talk with peers is the general interest in and curiosity about the topics discussed and finding them informative and/or intriguing. In that regard, obtaining (additional) information has been identified as a key motivation for reading users' comments (Diakopoulos & Naaman, 2011; Springer et al., 2015; Stroud et al., 2016), and that dynamic may also extend to uncivil discussions encompassing hate speech or contentious debates that are often viewed in a negative light. Research has indicated that negative events can have significant informational value, for confronting negativity plays a role in developing a more realistic understanding of the world (Baumeister et al., 2001). Likewise, although curiosity is typically associated with seeking positive stimuli and information, Oosterwijk (2017) discovered that morbid content is also subject to curiosity, such that individuals intentionally prefer to engage with images depicting negative social scenes (e.g., violent conflicts) instead of positive or neutral images. Information seeking and curiosity have also proven to be common reasons for users to watch even extreme content—for instance, ISIS beheading videos (Redmond et al., 2019).

However, not engaging with hate speech might also relate to cognitive motives, specifically the perceived lack of cognitive reward. In such cases, engaging with hate speech might simply be considered to be irrelevant and unnecessary due to not providing any informative value. Among internet users in Germany, for instance, 57% perceive hate speech comments as being uninteresting (Landesanstalt für Medien NRW, 2023), which might hinder their engagement. Such non-engagement corroborates recent findings indicating trends of the normalization of hate speech and social media users' desensitization to it, especially among young, frequent users (Schmid, Kümpel, & Rieger, 2024). Because hate speech and harassment are pervasive online, they are increasingly perceived by young social media users as a tolerable norm instead of serious incidents that need to be addressed (Haslop et al., 2021).

Altogether, the theoretical background on the use of media content with reference to the uses and gratifications approach suggests that a range of affective, cognitive, personal, and social motivations can impact individuals' attention to various types of media content, including negative content. Examining the current landscape of research addressing online hate speech reveals that similar factors

frequently influence how individuals perceive such content, their behavioral reactions to it, and the spread of hatred itself. Combining those strands of research, in our study we aimed to investigate whether social media users have similar motivations for their (non-)engagement with hate speech, conceptualized as the close examination, consideration, and willingness to reflect on the messages and their potential consequences (i.e., engagement) or the conscious decision to refrain from doing so (i.e., non-engagement).

Method

Design and Participants

In June and July 2022, we conducted a preregistered quantitative online survey of internet users in Germany aged at least 16 years old, recruited via respondi/Bilendi with quotas for age, gender, and federal state. Respondents participated voluntarily and were compensated in the company's internal credit point system. Considering our preregistered criteria for quality, we excluded 541 respondents (12%) who finished the survey too quickly based on the relative speed index (Leiner, 2019; see preregistration for details). As a result, the final sample comprised 4,020 participants (see Table 1).

Measures and Analyses

Our study was part of a larger preregistered project on social media users' engagement with hate speech and their intentions to intervene.[2] To save space, we report only the variables that are relevant to this article, with a focus on social media users' motives for and against engaging with hate speech.[3] Before examining participants' motives, we asked how often they had encountered hate speech on social media platforms in general on a 7-point scale ranging from 1 (*never*) to 7 (*very often; M*=3.37, *SD*=1.90). To ensure a common understanding, participants were given a definition of *hate speech*, with particular emphasis on its group-related nature and the variety of possible ways of expressing hate, including, for instance, hate speech targeting religious affiliation, origin, or sexual orientation in varying severity (see questionnaire in the OSF repository).

For motives of (non-)engagement with hate speech, participants who had witnessed hate speech on social media at least once (76%, *n*=3,051) rated 21 items regarding their engagement and 10 items regarding their non-engagement. In line with our theoretical framework, we developed items aligned with the proposed categories of affective and entertainment, personal identity and social-integrative as well as cognitive motives. It is important to note that this categorization is not exclusive, as some items cannot clearly be assigned to be just touching upon, for instance, affective or social integrative motives. To address our specific case of engaging and not engaging with hate speech, we developed concrete

Table 1. Sample Description.

Sociodemographic variables	N	%
Gender		
Male	2,045	51
Female	1,975	49
Age (Range: 16–92)		
16–29	729	18
30–39	566	14
40–49	640	16
50–59	880	22
50+	1,205	30
Educational level		
High	1,411	35
Middle	1,268	32
Low	1,341	33
Previous Exposure to Hate Speech		
No	969	24
Yes	3,051	76

Note. N=4,020, all participants after data cleaning. Educational level "high" includes general higher education entrance qualification / vocational baccalaureate and university degree, "middle" includes secondary school leaving certificates; "low" includes no general school leaving certificate and secondary / middle / elementary school leaving certificate.

wordings for items in an exploratory way but also adopted some items from past research on hate speech (Landesanstalt für Medien NRW, 2023), entertainment experiences (Oliver & Raney, 2011; translated by Odağ et al., 2018), and inspirational media use (Oliver et al., 2012).

Participants' agreement to the (non-)engagement items was measured on 7-point scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), following the introductory sentence, "When I encounter hate speech on social media platforms, I engage with it, . . .," respectively ". . . I don't engage with it, ..." Affective and entertainment motives were examined both for engagement (e.g., "Because hate speech worries me" or "Because I consider it entertaining") and non-engagement (e.g., "Because it is not good for me [e.g., my mental health]"). Regarding personal identity and social-integrative motives, we examined items representing social and collective identity (e.g., "To see that other people have the same opinion as I do"), social exchange (e.g., "To be able to talk about it with others"), and personal involvement in and perceived responsibility for both engagement (e.g., "If I know the attacked person[s]") and non-engagement (e.g., "Because I do not feel responsible for it"). Last, cognitive motives were assessed regarding engagement (e.g., "Because I consider it interesting") and non-engagement (e.g., "Because I do not consider it interesting"). The complete list of items for engagement appears in Table 2 and for non-engagement in Table 3.

Based on those measures, we conducted two exploratory factor analyses (i.e., method of minimal residuals) with oblique rotation (i.e., oblimin), as detailed in the study's preregistration.

Table 2. Rotated Factor Loadings for Exploratory Factor Analysis on Engaging With Hate Speech.

When I encounter hate speech on social media platforms, I engage with it, …	I PERSONAL INVOLVEMENT	II REFLECTION & INSPIRATION	III FUN & ENTERTAINMENT	IV SELF-AFFIRMATION	V PARTICIPATION IN DISCUSSIONS	VI APPALLED & CONCERNED
if I know the attacked person(s).	**0.93**	0.00	−0.06	−0.05	−0.01	−0.01
if I personally know the author(s) of the hate speech.	**0.86**	−0.03	0.05	0.02	−0.06	−0.06
if I feel I belong to the attacked group.	**0.68**	−0.03	0.00	0.09	0.10	0.08
if I know people who react to it (e.g., point to the hate speech or engage in counter speech).	**0.65**	0.10	0.07	0.03	0.05	0.06
if it concerns a topic that is personally relevant to me.	**0.56**	0.10	0.07	0.03	0.11	0.09
because it makes me think.	0.06	**0.79**	−0.07	−0.03	−0.01	0.09
because it challenges my way of seeing the world.	0.01	**0.76**	0.03	0.07	0.03	−0.02
because it motivates me to be a better person.	0.03	**0.55**	0.12	0.01	0.05	0.11
because it inspires me.	−0.02	**0.40**	0.19	0.19	0.21	−0.18
because I want to learn about the opinion of others.	0.06	**0.37**	0.18	0.15	0.18	−0.05
because I consider it entertaining.	0.00	0.02	**0.97**	0.00	−0.03	−0.02
because I enjoy observing arguments.	0.01	−0.03	**0.86**	0.01	0.05	0.03
to confirm me in my own similar opinion.	0.00	−0.01	0.00	**0.93**	−0.02	−0.01
to see that other people have the same opinion as I do.	0.01	0.01	−0.01	**0.89**	0.02	0.02
to be able to participate in the relevant discussion.	0.01	−0.05	0.00	0.02	**0.94**	−0.01
to be able to talk about it with others.	0.01	0.10	0.02	−0.02	**0.78**	0.03
because I am appalled by hate speech.	0.00	−0.02	0.03	−0.01	−0.02	**0.96**
because hate speech worries me.	0.01	0.05	−0.04	0.02	0.03	**0.86**
Generated Indices						
M (SD)	4.14 (1.62)	3.58 (1.44)	2.72 (1.75)	3.22 (1.77)	3.30 (1.75)	4.93 (1.82)
Cronbach's α / Spearman Brown	.88 .88	.83	.88	.88	.85	.89

Note. $n = 3,051$; Participants who have seen hate speech before; KMO = .93 (KMO-values for single items ≥ .80); Bartlett-test:($\chi^2(210) = 67,268.23$, $p < .001$; Parallel analysis suggested six-factor solution; factor analysis (method of minimal residuals) with oblique rotation (Oblimin) was performed for the total of 21 items (scale: 1 [*strongly disagree*] to 7 [*strongly agree*]). Three items ("because I consider it interesting"; "in order to then possibly do something against the hate speech," "because I think it is necessary to deal with it") were excluded due to cross-loadings in a second and third analysis step. The final EFA reported here was thus only conducted with 18 items. All factor loadings >.30 are in bold. The items marked as such in the respective columns were combined to mean indices.

Table 3. Rotated Factor Loadings for Exploratory Factor Analysis on Not Engaging With Hate Speech.

When I encounter hate speech on social media platforms, I don't engage with it, . . .	I	II	III	IV
	IRRELEVANCE	NO RESPONSIBILITY	SELF-PROTECTION	CONSCIOUS IGNORING
because it is not worth the effort for me.	**0.87**	0.01	−0.04	0.04
because I do not believe that it brings (me) anything.	**0.84**	0.00	0.06	−0.02
because I do not consider it interesting.	**0.36**	0.22	0.06	0.14
because I do not think it is necessary to deal with it.	−0.01	**0.91**	0.00	0.03
because I do not feel responsible for it.	0.19	**0.59**	0.10	−0.06
because it is not good for me (e.g., my mental health).	0.03	−0.07	**0.82**	−0.03
to avoid the hate.	−0.02	0.08	**0.81**	0.05
because I do not want to give attention to the authors of hate speech.	0.02	−0.04	0.02	**0.88**
because I think that conscious ignoring is more effective.	0.04	0.25	0.05	**0.53**
Generated Indices				
M (SD)	4.53 (1.66)	4.09 (1.74)	4.47 (1.80)	4.88 (1.67)
Cronbach's α / Spearman Brown	.80	.77	.76	.70

Note. n = 3,051; Participants who have seen hate speech before; KMO = .90 (KMO-values for single items ≥ .83); Bartlett-test:(χ^2(45) = 21,726.09, *p* < .001; Parallel analysis suggested five-factor solution; factor analysis (method of minimal residuals) with oblique rotation (Oblimin) was performed for the total of 10 items (scale: 1 [*strongly disagree*] to 7 [*strongly agree*]). One item ("because I am not interested in the opinion of others") was excluded in a second analysis step. The final EFA reported here was thus only conducted with nine items, with the respective parallel analysis suggesting the reported four-factor solution. All factor loadings > .30 are in bold. The items marked as such in the respective columns were combined to mean indices.

For statistical parameters, see Tables 2 and 3. Ultimately, we extracted six motives for social media users' engagement with hate speech and four motives for their non-engagement.

Results

Motives for Engaging With Hate Speech

We conducted a parallel analysis to determine the number of factors to retain for the engagement with hate speech. The results suggested the extraction of six factors—that is, motives—as detailed in Table 2. One was *Personal Involvement* (*M* = 4.14, *SD* = 1.62, α = .88), meaning motives related to being a member of the target group, knowing the author and/or people responding to the hate speech, and/or knowing the individuals affected. Another was *Reflection and Inspiration* (*M* = 3.58, *SD* = 1.44, α = .83), meaning motives related to learning from others' perspectives, being inspired and/or challenged in relation to one's worldview, and/or being motivated to be a better person. The third motive was *Fun and Entertainment* (*M* = 2.72, *SD* = 1.75, α = .88), meaning motives related to the enjoyment of observing arguments and being entertained, while the fourth was *Self-Affirmation* (*M* = 3.22, *SD* = 1.77, α = .88), meaning motives related to being affirmed in one's opinion and perceiving that others share it. The fifth was *Participation in Discussions* (*M* = 3.30, *SD* = 1.75, α = .85), meaning motives related to being able to participate in a discussion and talk about the topic. The sixth and final motive was *Appalled and Concerned* (*M* = 4.93, *SD* = 1.82, α = .89), meaning motives for

engagement related to being appalled by and worried about hate speech.

Motives for Not Engaging With Hate Speech

For non-engagement with hate speech, the parallel analysis suggested a theoretically plausible and unambiguous four-factor solution, as detailed in Table 3. The first was *Irrelevance* (*M* = 4.53, *SD* = 1.66, α = .80), meaning motives related to the belief that engagement is pointless, uninteresting, and/or not worth the effort. The second was *No Responsibility* (*M* = 4.09, *SD* = 1.74, α = .77), meaning motives related to perceiving engagement as being unnecessary or not considering oneself to be responsible, while the third was *Self-Protection* (*M* = 4.47, *SD* = 1.80, α = .76), meaning motives related to protecting oneself from negative effects. The fourth and final factor was *Conscious Ignoring* (*M* = 4.88, *SD* = 1.67, α = .70), meaning motives related to believing that deliberately ignoring hate speech is more efficient than giving attention to its perpetrators.

Discussion

Despite considerable research on what causes social media users to recognize hate speech (i.e., first step of recognition) and what motivates their reactions toward it (i.e., third step of behavior), little is known about an important intermediary step, namely, users' engagement or non-engagement with the content (see Figure 1). We have argued that such an intermediary step is worth investigating, however, because it provides the cognitive preconditions for subsequent reactions to

hate speech, whether in condemnation or endorsement. Furthermore, analyzing users' motives for their engagement provides crucial insights for preventive work and for developing effective strategies to combat hate speech. As we assumed, various factors contribute to whether users engage with hate speech or consciously abstain from doing so. Grounded in the uses and gratifications approach, we examined affective and entertainment motives, personal identity and social-integrative motives, and cognitive motives for engaging or not engaging with hate speech online.

Summary of Findings and Implications

The most important reasons for users to engage with content expressing hate speech relate to negative affect due to being shocked and worried about it (i.e., *Appalled and Concerned*) and personally implicated (i.e., *Personal Involvement*), both of which have been shown to elicit attention to hate speech—that is, recognition of it (Schmid, Kümpel, & Rieger, 2024)—and to influence users' decision to intervene, meaning their behavior (Gagrčin, 2022). In addition, important but somewhat less pronounced was being inspired to think more about the topic or challenge one's opinions (i.e., *Inspiration and Reflection*), a finding that contributes to research on eudaimonic entertainment by highlighting its significance in explaining people's attraction to online hate speech, similar to their attraction to violent media content (e.g., Bartsch et al., 2016). That result was somewhat expected given that extremist communicators, who often disseminate hateful content, frequently incorporate elements of eudaimonic entertainment into their content (Frischlich, 2021). Our analysis also suggests that users may turn to hate speech to participate in relevant discussions (i.e., *Participation in Discussions*). That motive of social connectivity was previously found for reading users' comments in general (Diakopoulos & Naaman, 2011; Springer et al., 2015; Stroud et al., 2016) and can now be transferred to negative content. Although relatively few participants reported engaging with hate speech because it supports their opinion (i.e., *Self-Affirmation*) or amuses them (i.e., *Fun and Entertainment*), those motives should not be overlooked because they are corroborated by past findings identifying positive in-group emotions (Walther, 2022), fun (Erjavec & Kovačič, 2012), enthusiasm (Kosmidis & Theocharis, 2020; Sydnor, 2018), and sensation seeking (Bedrosova et al., 2023) as motivators for posting or intentionally encountering hate speech. The finding that hate speech is also consumed for entertainment somehow speaks to the idea that toxic personality traits of (very few) users are responsible for toxic online communication. Indeed, a recent study has shown that toxic users polluted communication on Reddit in both partisan and nonpartisan contexts (Mamakos & Finkel, 2023). Likewise, several studies have provided evidence of an association between the "Dark Triad" of personality traits

and posting hate comments online (Frischlich et al., 2021; Sorokowski et al., 2020). We further assume that the motive to engage with hate speech for the sake of entertainment, like any other motive, is not only shaped by personality traits and attitudes toward hate speech but also by certain types of speech, specifically hate speech containing humorous elements (Schmid, 2023).

Overall, our analysis revealed that negative affect, particularly in the form of concern, was the most prominent motivation for engaging with hate speech, often linked to users' rejection of the content. Conversely, motivations rooted in personal approval and entertainment were relatively subdued in our sample, whereas reasons such as participation in discussions and seeking reflection and inspiration fell somewhere in between. Those findings suggest that engagement with hate speech provides social media users with a spectrum of personal and social gratifications that extend beyond mere expressions of approval or disapproval.

Regarding non-engagement with hate speech, consciously ignoring it appeared to be the primary reason, one driven by the notion that confronting haters only amplifies their visibility (i.e., *Conscious Ignoring*). However, we also found that dealing with hate speech is sometimes perceived to be pointless because intervening has no reward and/or is simply not worth the effort (i.e., *Irrelevance*). The perception of its irrelevance also relates to some social media users' perceived lack of responsibility (i.e., *No Responsibility*), which, along with the belief that engaging with hate speech is pointless, may contribute to its normalization, such that it is acknowledged but not regarded as a (personally) relevant problem (Bilewicz & Soral, 2020; Cover, 2023; Haslop et al., 2021). Last, avoiding hate and considering engagement with hate speech to be unhealthy (i.e., *Self-Protection*) emerged as a reasonable motive for some social media users, especially given its potential threats to personal mental health (Cover, 2022; Gelber & McNamara, 2016). Such a self-protective motivation supports recent findings showing that some social media users purposefully refrain from engaging with hate speech and other potentially uncivil (political) debates after experiencing unpleasant encounters with hate speech (Barnidge et al., 2019; Gagrčin & Milzner, 2023).

Overall, the mean agreement for the motives related to non-engagement slightly exceeded those related to engagement. Although we cannot draw any conclusions about the actual extent of (non-)engagement, that outcome confirms our assumptions that deliberately avoiding hate speech is appropriate for some social media users, at least in certain situations and for certain types of hate speech. However, according to our data, if action is not taken to engage with or combat hate speech online, then it would be incorrect to believe that people are generally disinterested in or do not feel responsible for intervening. Instead, there are far more nuanced explanations for their non-engagement, some of which are undoubtedly situational and rely on a motivation to shield oneself or others from harm. We, therefore, encourage

research that more closely examines the implications of non-engagement and considers what it means for practical work related to preventing the perpetration of hate speech or mitigating its negative personal and societal effects.

Limitations and Directions for Future Research

In building our study on the uses and gratifications approach, we were able to identify and discuss social media users' various motives for and against engaging with hate speech. Following that approach, we considered (non-)engagement to result from a conscious decision, one that individuals possess the capacity to articulate and communicate. In doing so, we ran the risk of disregarding more unconscious forms of engagement or those that have become ingrained as habitual behaviors, thereby making them challenging to recognize or reflect upon. That constraint relates to a more general debate about the uses and gratifications approach and the common critique of assuming an actively acting and deciding audience (Ruggiero, 2000). Regarding cognitive dissonance theory (Festinger, 1957), for instance, predisposed selective attention to hate speech could exist that influences users' subsequent levels of (non-)engagement. Considering that research on hate speech has mostly neglected motives for non-engagement, it is probable that our survey did not encompass all possible motives. To thoroughly explore those motives, subsequent research could employ qualitative interviews or observational methods that allow differentiating various scenarios involving hate speech and thereby elucidating diverse motivational dynamics among social media users. Building on our work, future research should also more thoroughly consider the specific content and targets of hate speech, the (social media) context, and recipients' personal characteristics, all of which have been shown to influence perceptions of hate speech and behavioral reactions to it. For example, recent research emphasizes the perceptual nature of hate speech and incivility by demonstrating that different types of norm violations vary in how offensive and harmful they are perceived (Kümpel & Unkel, 2023). Furthermore, people's expectations regarding the outcomes of their (non-)engagement may differ greatly. Understanding these aspects by accounting for further content- and personal-related characteristics in follow-up research is particularly important, because users engage with hate speech for a multitude of (potentially conflicting) reasons depending on the context, and the degree of their engagement can vary significantly.

Follow-up research is also essential to elucidating the complex, possibly reciprocal relationships between users' motives for (not) engaging with hate speech, their actual experiences during such engagement, and the lessons that they draw from it. In this article, we have conceptualized a process of reception in which users' motives precede their actual engagement with hate speech. However, the impact of emotions that surface only during deep engagement—for instance, amusement and anxiety—remains incompletely understood. On that

count, feedback loops related to gratifications obtained from prior encounters with hate speech likely significantly influence users' subsequent decisions to (not) engage.

Last, our findings are limited to the German context, which has rather strict anti-hate speech legislation compared with, for instance, the United States (Hawdon et al., 2017). Examining international differences in motives for (non-)engagement based on varying cultural, historical, and legal contexts could thus be a valuable avenue for future research as well.

ORCID iDs

Ursula Kristin Schmid 🔟 https://orcid.org/0000-0002-1892-002X
Anna Sophie Kümpel 🔟 https://orcid.org/0000-0001-7184-4057
Diana Rieger 🔟 https://orcid.org/0000-0002-2417-0480

Notes

1. By using the term *engagement with hate speech*, we do not mean personally using hate speech or offensive language although such behavior might result from engaging with the hate speech of others.
2. The preregistration for the whole project can be viewed at https://osf.io/8qx9n
3. A translated version of the questionnaire and a full description of its items can be found in our OSF repository at https://osf.io/kv5t7

References

Barnidge, M., Kim, B., Sherrill, L. A., Luknar, Ž., & Zhang, J. (2019). Perceived exposure to and avoidance of hate speech in various communication settings. *Telematics and Informatics*, *44*, 101263. https://doi.org/10.1016/j.tele.2019.101263

Bartsch, A., Mares, M.-L., Scherr, S., Kloß, A., Keppeler, J., & Posthumus, L. (2016). More than shoot-em-up and torture porn: Reflective appropriation and meaning-making of violent media content. *Journal of Communication*, *66*(5), 741–765. https://doi.org/10.1111/jcom.12248

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bedrosova, M., Mýlek, V., Dedkova, L., & Velicu, A. (2023). Who is searching for cyberhate? Adolescents' characteristics associated with intentional or unintentional exposure to cyberhate. *Cyberpsychology, Behavior and Social Networking*, *26*, 462–471. https://doi.org/10.1089/cyber.2022.0201

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*(Suppl. 1), 3–33. https://doi.org/10.1111/pops.12670

Bormann, M., Tranow, U., Vowe, G., & Ziegele, M. (2021). 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

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Cover, R. (2022). Digital hostility: Contemporary crisis, disrupted belonging and self-care practices. *Media International Australia*, *184*(1), 79–91. https://doi.org/10.1177/1329878X221088048

Cover, R. (2023). Digital hostility, subjectivity and ethics: Theorising the disruption of identity in instances of mass online abuse and hate speech. *Convergence*, *29*(2), 308–321. https://doi.org/10.1177/13548565221122908

Diakopoulos, N., & Naaman, M. (2011). Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 133–142). Association for Computing Machinery. https://doi.org/10.1145/1958824.1958844

Elnakouri, A., Hubley, C., & McGregor, I. (2022). Hate and meaning in life: How collective, but not personal, hate quells threat and spurs meaning in life. *Journal of Experimental Social Psychology*, *98*, 104227. https://doi.org/10.1016/j.jesp.2021.104227

Erjavec, K., & Kovačič, M. P. (2012). "You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. *Mass Communication and Society*, *15*(6), 899–920. https://doi.org/10.1080/15205436.2011.619679

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Frischlich, L. (2021). #Dark inspiration: Eudaimonic entertainment in extremist Instagram posts. *New Media & Society*, *23*(3), 554–577. https://doi.org/10.1177/1461444819899625

Frischlich, L., Schatto-Eckrodt, T., Boberg, S., & Wintterlin, F. (2021). Roots of incivility: How personality, media use, and online experiences shape uncivil participation. *Media and Communication*, *9*(1), Article 1. https://doi.org/10.17645/mac.v9i1.3360

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*, *26*, 4299–4316. https://doi.org/10.1177/14614448221117499

Gagrčin, E., & Milzner, M. (2023). "Intervening is a good thing but . . .": The role of social norms in users' justifications of (non-) intervention against incivility. *Social Media + Society*, *9*(3). https://doi.org/10.1177/20563051231186561

Gelber, K., & McNamara, L. (2016). Evidencing the harms of hate speech. *Social Identities*, *22*(3), 324–341. https://doi.org/10.1080/13504630.2015.1128810

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, *12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Grizzard, M., Huang, J., Weiss, J. K., Novotny, E. R., Fitzgerald, K. S., Ahn, C., Ngoh, Z., Plante, A., & Chu, H. (2017). Graphic violence as moral motivator: The effects of graphically violent content in news. *Mass Communication and Society*, *20*(6), 763–783. https://doi.org/10.1080/15205436.2017.1339804

Haslop, C., O'Rourke, F., & Southern, R. (2021). #NoSnowflakes: The toleration of harassment and an emergent gender-related digital divide, in a UK student online culture. *Convergence*, *27*(5), 1418–1438. https://doi.org/10.1177/1354856521989270

Hawdon, J., Oksanen, A., & Räsänen, P. (2017). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior*, *38*(3), 254–266. https://doi.org/10.1080/01639625.2016.1196985

Hu, Y., Zhang, T., Shi, H., & Fan, C. (2023). Empathy and bystander helping behavior in cyberbullying among adolescents: The mediating role of internet moral judgment and the moderating role of internet self-efficacy. *Frontiers in Psychology*, *14*, Article 1196571. https://doi.org/10.3389/fpsyg.2023.1196571

Jost, P., Ziegele, M., & Naab, T. K. (2020). Klicken oder tippen? Eine Analyse verschiedener Interventionsstrategien in unzivilen Online-Diskussionen auf Facebook [Click or type? An analysis of different intervention strategies in uncivil online discussions on Facebook]. *Zeitschrift für Politikwissenschaft*, *30*(2), 193–217. https://doi.org/10.1007/s41358-020-00212-9

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *The Public Opinion Quarterly*, *37*(4), 509–523.

Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, *66*, 236–247. https://doi.org/10.1016/j.chb.2016.09.024

Kosmidis, S., & Theocharis, Y. (2020). Can social media incivility induce enthusiasm? *Public Opinion Quarterly*, *84*(S1), 284–308. https://doi.org/10.1093/poq/nfaa014

Kümpel, A. S., & Unkel, J. (2023). Differential perceptions of and reactions to incivil and intolerant user comments. *Journal of Computer-Mediated Communication*, *28*(4), zmad018. https://doi.org/10.1093/jcmc/zmad018

Landesanstalt für Medien NRW. (2023). *Hate Speech Forsa-Studie 2023. Zentrale Untersuchungsergebnisse*. https://www.medienanstalt-nrw.de/themen/hass/forsa-befragung-zur-wahrnehmung-von-hassrede.html

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in Internet surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Leonhard, L., Rueß, C., Obermaier, M., & Reinemann, C. (2018). Perceiving threat and feeling responsible. How severity of hate speech, number of bystanders, and prior reactions of others affect bystanders' intention to counterargue against hate speech on Facebook. *Studies in Communication and Media*, *7*(4), 555–579. https://doi.org/10.5771/2192-4007-2018-4-555

Mamakos, M., & Finkel, E. J. (2023). The social media discourse of engaged partisans is toxic even when politics are irrelevant. *PNAS Nexus*, *2*(10), pgad325. https://doi.org/10.1093/pnasnexus/pgad325

Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, *20*(6), 930–946. https://doi.org/10.1080/1369118X.2017.1293130

Matamoros-Fernández, A. (2018). Inciting anger through Facebook reactions in Belgium: The use of emoji and related vernacular expressions in racist discourse. *First Monday*, *23*(9). https://doi.org/10.5210/fm.v23i9.9405

McQuail, D. (1987). *Mass communication theory: An introduction* (2nd ed.). Sage.

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, *99*(1), 1–15. https://doi.org/10.1017/S0003055405051452

Niehoff, E., & Oosterwijk, S. (2020). To know, to feel, to share? Exploring the motives that drive curiosity for negative content. *Current Opinion in Behavioral Sciences*, 35, 56–61. https://doi.org/10.1016/j.cobeha.2020.07.012

Odağ, Ö., Uluğ, Ö. M., Arslan, H., & Schiefer, D. (2018). Culture and media entertainment: A cross-cultural exploration of hedonic and eudaimonic entertainment motivations. *International Communication Gazette*, *80*(7), 637–657. https://doi.org/10.1177/1748048518802215

Oliver, M. B., Hartmann, T., & Woolley, J. K. (2012). Elevation in response to entertainment portrayals of moral virtue. *Human Communication Research*, *38*(3), 360–378. https://doi.org/10.1111/j.1468-2958.2012.01427.x

Oliver, M. B., & Raney, A. A. (2011). Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication*, *61*(5), 984–1004. https://doi.org/10.1111/j.1460-2466.2011.01585.x

Oosterwijk, S. (2017). Choosing the negative: A behavioral demonstration of morbid curiosity. *PLOS ONE*, *12*(7), Article e0178399. https://doi.org/10.1371/journal.pone.0178399

Redmond, S., Jones, N. M., Holman, E. A., & Silver, R. C. (2019). Who watches an ISIS beheading-And why. *The American Psychologist*, *74*(5), 555–568. https://doi.org/10.1037/amp0000438

Rieger, D., Frischlich, L., & Bente, G. (2020). Dealing with the dark side: The effects of right-wing extremist and Islamist extremist propaganda from a social identity perspective. *Media, War & Conflict*, *13*(3), 280–299. https://doi.org/10.1177/1750635219829165

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

Rozin, P., Guillot, L., Fincher, K., Rozin, A., & Tsukayama, E. (2013). Glad to be sad, and other examples of benign masochism. *Judgment and Decision Making*, *8*(4), 439–447. https://doi.org/10.1017/S1930297500005295

Ruggiero, T. E. (2000). Uses and gratifications theory in the 21st century. *Mass Communication and Society*, *3*(1), 3–37. https://doi.org/10.1207/S15327825MCS0301_02

Schäfer, S., Rebasso, I., Boyer, M. M., & Planitzer, A. M. (2023). Can we counteract hate? Effects of online hate speech and counter speech on the perception of social groups. *Communication Research*, *51*, 553–579. https://doi.org/10.1177/00936502231201091

Schmid, U. K. (2023). Humorous hate speech on social media: A mixed-methods investigation of users' perceptions and processing of hateful memes. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448231198169

Schmid, U. K., Kümpel, A. S., & Rieger, D. (2024). How social media users perceive different forms of online hate speech: A qualitative multi-method study. *New Media & Society*, *26*(5), 2614–2632. https://doi.org/10.1177/14614448221091185

Schmid, U. K., Obermaier, M., & Rieger, D. (2024). Who cares? How personal political characteristics are related to online counteractions against hate speech. *Human Communication Research*, *50*(3), 393–403. https://doi.org/10.1093/hcr/hqae004

Sorokowski, P., Kowal, M., Zdybek, P., & Oleszkiewicz, A. (2020). Are online haters psychopaths? Psychological predictors of online hating behavior. *Frontiers in Psychology*, 11, Article 553. https://www.frontiersin.org/article/10.3389/fpsyg.2020.00553

Springer, N. (2014). *Beschmutzte Öffentlichkeit? Warum Menschen die Kommentarfunktion auf Online-Nachrichtenseiten als öffentliche Toilettenwand benutzen, warum Besucher ihre Hinterlassenschaften trotzdem lesen, und wie die Wände im Anschluss aussehen* [Soiled public? Why people use the comment function on online news sites as a public toilet wall, why visitors still read what they leave behind, and what the walls look like afterwards]. LIT Verlag Münster.

Springer, N., Engelmann, I., & Pfaffinger, C. (2015). User comments: Motives and inhibitors to write and read. *Information, Communication & Society*, *18*(7), 798–815. https://doi.org/10.1080/1369118X.2014.997268

Stroud, N. J., Duyn, E. V., & Peacock, C. (2016). *News commenters and news comment readers*. Center for Media Engagement. https://mediaengagement.org/research/survey-of-commenters-and-comment-readers/

Sydnor, E. (2018). Platforms for incivility: Examining perceptions across different media formats. *Political Communication*, *35*(1), 97–116. https://doi.org/10.1080/10584609.2017.1355857

Walther, J. B. (2022). Social media and online hate. *Current Opinion in Psychology*, *45*, 101298. https://doi.org/10.1016/j.copsyc.2021.12.010

Williams, A., Oliver, C., Aumer, K., & Meyers, C. (2016). Racial microaggressions and perceptions of Internet memes. *Computers in Human Behavior*, *63*, 424–432. https://doi.org/10.1016/j.chb.2016.05.067

Zerback, T., & Fawzi, N. (2017). Can online exemplars trigger a spiral of silence? Examining the effects of exemplar opinions on perceptions of public opinion and speaking out. *New Media & Society*, *19*(7), 1034–1051. https://doi.org/10.1177/1461444815625942

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

Ursula Kristin Schmid (MA, LMU Munich) is a Research Associate in the Department of Media and Communication at LMU Munich, Germany. Her research focuses the characteristics, perceptions and effects of hate speech, online incivility and counter speech on social media. In her PhD project, she investigates the interplay of online hate speech and humor.

Anna Sophie Kümpel (PhD, LMU Munich) is a Full Professor of Communication in the Department of Media and Communication at LMU Munich, Germany. Her research explores the dynamics of (political) information use in digital and algorithmically driven environments, particularly within social media. She investigates how these platforms shape user engagement, perceptions, and information dissemination.

Diana Rieger (PhD, University of Cologne) is a Full Professor at the Department of Media and Communication at LMU Munich, Germany. Her current work addresses characteristics and effects of hate speech, extremist online communication and counter voices (e.g., counter narratives and counter speech). Her second focus is on entertainment research, investigating how entertaining online content (e.g. memes, TikToks) is processed and experienced.