UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Uncivil user comments increase users'intention to engage in corrective actions and their support for authoritative restrictive actions

Teresa K. Naab, Thorsten Naab, Jonas Brandmeier

Angaben zur Veröffentlichung / Publication details:

Naab, Teresa K., Thorsten Naab, and Jonas Brandmeier. 2021. "Uncivil user comments increase users'intention to engage in corrective actions and their support for authoritative restrictive actions." *Journalism & Mass Communication Quarterly* 98 (2): 566–88. https://doi.org/10.1177/1077699019886586.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

Uncivil User Comments Increase Users' Intention to Engage in Corrective Actions and Their Support for Authoritative Restrictive Actions

Teresa K. Naab, Thorsten Naab, & Jonas Brandmeier

The published version of this manuscript is available here:

Naab, T. K., Naab, T. & Brandmeier, J. (2019). Uncivil user comments increase users' intention to engage in corrective actions and their support for authoritative restrictive actions. Journalism & Mass Communication Quarterly. Advance online publication. https://doi.org/10.1177/1077699019886586

Dr. Teresa K. Naab, Department of Media, Knowledge and Communication, University of Augsburg, Universitaetsstrasse 10, 86159 Augsburg, Germany. E-Mail: teresa.naab@phil.uni-augsburg.de, phone: +498215985933, https://orcid.org/0000-0001-7345-2559. Twitter: @NaabTeresa. Ph.D. granting institution: Hanover University of Music, Drama and Media, Germany.

Dr. Thorsten Naab, German Youth Institute, Nockherstrasse 2, 81541 München. E-Mail: naab@dji.de, phone: +49 89 62306-558. Ph.D. granting institution: Hanover University of Music, Drama and Media, Germany.

Jonas Brandmeier, BA, Seminar for Media and Communication Studies, University of Erfurt, Nordhaeuser Strasse 63, 99089 Erfurt, Germany. E-Mail: jonas.brandmeier@web.de.

Institution granting bachelor's degree: University of Augsburg, Germany.

Correspondence concerning this article should be addressed to Dr. Teresa K. Naab, Department of Media, Knowledge and Communication, University of Augsburg,

Universitaetsstrasse 10, 86159 Augsburg, Germany. E-Mail: teresa.naab@phil.uni-augsburg.de, phone: +498215985933.

Abstract

In an experimental study (n = 213) we vary the incivility of a user comment below a mock news post on the Facebook page of a German news outlet. We investigate determinants of the users' intention to engage in corrective actions and their support for restrictions by Facebook and law enforcement. We test an integrated model and find that incivility of a comment affects users' support for restrictive and intention for corrective actions, while the presumed influence of the comment does not have an effect. The users' commitment to free speech decreases and paternalism increases their support for authoritative restrictions.

Keywords: corrective and restrictive actions, presumed influence, user comments, incivility

Uncivil User Comments Increase Users' Intention to Engage in Corrective Actions and Their Support for Authoritative Restrictive Actions

Comment sections on news outlets' social media pages provide a public forum for users to express their opinions on current issues (Dahlberg, 2011). Despite the valuable contribution of online discussions to the public sphere (Graham & Wright, 2015), a relevant share of comments is uncivil and perceived as undesirable by readers (Duggan & Smith, 2016) and news outlets (Stroud, Scacco, Muddiman, & Curry, 2015). Uncivil comments "convey an unnecessarily disrespectful tone toward the discussion forum, its participants, or its topics" (Coe et al., 2014, p. 660). Following Coe and colleagues (2014), they include various forms like name-calling, aspersion, lying, vulgarity, and pejorative speech. Papacharissi (2004) names these aspects impoliteness and adds that incivility can include even more extreme forms of intolerant speech such as stereotypes against social groups, refusing equal rights to all, or threats against democracy. In line with this, other scholars have also complemented the list of uncivil expressions with various forms of undemocratic speech, such as racist or sexist terms (Chen, 2017). Under the term incivility we will consider all of these aspects because they can all threaten the quality of conversations and reach beyond sole negativity (Gervais, 2017).

The regulation of uncivil comments is under much public and scholarly debate. News outlets engage in the moderation of the seemingly endless amount of comments (e.g., Stroud et al., 2015), some have even shut down their comments sections (Thurman, Cornia, & Kunert, 2016) or transferred them to social network sites. Social media platforms have created extensive community standards along which they regulate content. In Germany, comments in online discussions are also under legislative regulation to ban anti-democratic misuse of freedom of expression (Leisegang, 2017). Our paper complements the debate by investigating the users' perspective on regulating uncivil comments on news pages. This

seems particularly relevant since the users command increased opportunities for corrective actions against disliked content and thus can complement regulation by media companies and the state.

We root our analysis in research on presumed influence (e.g., Rojas, 2010; Sun, Shen, & Pan, 2008). This research outlines that – as a consequence of perceiving media content as undesirable or influential – media users may wish to restrict or correct this content. To advance existing research, we reflect the opportunities of corrective actions that users experience in social media and the restrictive actions that they can expect from authorities – in particular from the platform provider Facebook and state regulation. We further aim at understanding underlying reasons for users' support for restrictions and their correction intentions. In an experimental design, we present participants (n = 213) with comments below a mock news post that was said to originate from the Facebook page of the German news outlet *SpiegelOnline*. The comments vary in that they express their opinion on the topic of the news post either civilly or uncivilly. In an integrated structural equation model, we test how the variation of incivility influences the users' perception of the comment's effect on themselves and others as well as their support for restrictive actions and their correction intentions. Moreover, this study is among the first to test how paternalistic tendency and the appreciation of freedom of expression influence the support for restrictive actions and the intention for corrective actions. The results provide insights into the determinants of user participation in the regulation of comment sections and the assessment of the regulation practices of platform providers and the state.

Restrictive and Corrective Actions

When confronted with media messages users perceive as undesirable or hostile against their view, they could engage in restrictive and corrective actions to limit these messages and their potential influence (Feng & Guo, 2012; Sun, Shen et al., 2008; Xu &

Gonzenbach, 2008). With restrictive actions, people seek to stop disliked content (Rojas, 2010). However, especially in traditional media, recipients have little possibilities of producing or interfering with existing content (for exceptions see Chia, Lu, & McLeod, 2004; Hoffner & Buchanan, 2002; Tsfati, Ribak, & Cohen, 2005). Thus, research on restrictive actions typically refers to people's support for restrictions by authorities who command greater power over content. For example, scholars investigate people's approval of government regulation (Dohle & Bernhard, 2014; Golan & Lim, 2016; Huh, Delorme, & Reid, 2006; Lim, 2017; Neuwirth & Frederick, 2002; Rojas, Shah, & Faber, 1996) or their support for restrictive actions by media organizations (McLeod, Detenber, & Eveland, 2001; McLeod, Eveland, & Nathanson, 1997; Wei & Lo, 2007).

"Corrective actions, in contrast to restrictive actions such as censorship or regulation, refer to individuals' engagement in reactive action against potentially harmful influence" (Lim, 2017, p. 978). They aim to counterbalance a message by correcting or contradicting the perceived wrong (Barnidge & Rojas, 2014; Rojas, 2010). The forms of corrective actions are heterogeneous. Considering online corrective action, researchers have investigated sharing countering information, engaging in online discussions, posting online comments, stating disapproval with a message in a reply comment or with Likes or Dislikes, and producing a countering video (Bernhard & Dohle, 2015; Chung, Munno, & Moritz, 2015; Golan & Lim, 2016; Hwang et al., 2008; Kalch & Naab, 2017; Lim, 2017; Lim & Golan, 2011; Naab, Kalch, & Meitz, 2018; Rojas, 2010). In sum, these examples point to the fact that social media empower ordinary users to counterbalance content they perceive in conflict with their values or social norms often bypassing elites (Lim & Golan, 2011).

The perceived incivility of a user comment and the perception that it will have negative effects (third-person perception) should influence users' attitude toward restrictive and corrective actions. We will outline both proposed influences as well as the

interconnections between the two. We will consider the tendency to paternalism as a central theoretical explanation. To more fully understand the mechanisms, we will juxtapose users' paternalism with their commitment to freedom of expression. This will help to provide a more elaborated picture of the contrary tendencies that may influence support for restrictive and intention to correction actions.

Influence of Incivility

Incivility challenges conversational norms and is perceived as socially deviant (Brown & Levinson, 1987; Grice, 1969; Stryker, Conway, & Danielson, 2016). It threatens the sense of face of the interaction partners. The face is the socially constructed identity that people act out during interactions (Goffman, 1967; Oetzel, Ting-Toomey, Yokochi, Masumoto, & Takai, 2000). When an interaction does not adequately meet people's face wants, people experience negative emotions (Brett et al., 2007). In line with this, uncivil comments may be perceived as norm violations (Muddiman, 2017) and face threat because they question the readers' need for autonomy and make them feel disrespected (Chen & Ng, 2016) and, thus, can cause negative feelings (Gervais, 2017; Rösner, Winter, & Krämer, 2016). Face theory further suggests that face-threatening acts may prompt retaliatory behaviors to restore one's face or hurt the face of the perpetrator in return (Brett et al., 2007). Online incivility has also been shown to prompt retaliatory responses (Chen & Lu, 2017). Thus, we assume that the perceived incivility of a comment (independently of the presumed susceptibility of an audience to its influence, see below) might have a direct influence on support for restrictions and engagement in corrective actions. These responses should arise from the recipients' negative emotions about the message or their desire to punish the communicator for the message (Chia et al., 2004; McLeod et al., 2001). A direct influence of the assessed severity of a message (in other words: incivility) on restrictive actions has been supported in numerous communication studies (McLeod et al., 2001; Paek, Lambe, &

McLeod, 2008; Shah, Faber, & Youn, 1999). In line with this, scholarship on hostile media perception indicates that when people presume that media coverage is negatively biased against their perspective they are inclined to restrictions and corrective actions (e.g., Feldman, Hart, Leiserowitz, Maibach, & Roser-Renouf, 2017; Hwang et al., 2008; Rojas, 2010; Wei et al., 2011).

To our knowledge, there is little counter-evidence to the influence of perceived severity on restrictive and corrective actions. However, Chia and colleagues (2004) do not find support for their hypothesis that perceived severity of an online video should increase support for censorship and reluctance to disseminate the video personally. Lim (2017) finds an influence of perceived undesirability on support for restrictions, but only an indirect effect of perceived undesirability on corrective actions via support for restrictions.

Concerning user comments, the overall picture of the findings suggests, that an uncivil user comment should increase support for restrictive and corrective actions compared to a civil comment:

H1: The support for restrictions by authorities is greater for an uncivil than a civil comment.

H2: The intention to engage in corrective actions is greater for an uncivil than a civil comment.

Influence of Third-Person Perception and Paternalism

Besides the perceived incivility of a user comment, the presumed influence of the comment should determine the users' support for restrictive actions and their intention to engage in corrective actions. Presumed influence of messages is examined under the third-person perception hypothesis. The third-person perception is the tendency for people to believe that others are more affected by the media than they are themselves (Davison, 1983). It can be explained by people's self-enhancement motivation: people perceive themselves as

superior to others in the sense of being less susceptible to potential negative influences. This discrepancy between presumed influences on self and others is a robust finding in communication research across a multitude of message domains (Paul, Salwen, & Dupagne, 2000; Sun, Pan, & Shen, 2008). It has also been supported for the presumed effects of user comments (Chen & Ng, 2016, 2017; Houston, Hansen, & Nisbett, 2011).

The perceptual discrepancy between presumed influence on self and others can lead to attitudinal and behavioral consequences, the so-called third-person effect. It is among the well-supported consequences of third-person perception that presumably influential media messages can prompt people to support restrictive actions to protect the vulnerable others from harmful effects (Feng & Guo, 2012; Xu & Gonzenbach, 2008). This holds for traditional media content (Bernhard & Dohle, 2014; Rosenthal, Detenber, & Rojas, 2018; Wei et al., 2011) as well as online content like online ads for cosmetic surgery (Lim, 2017), internet pornography (Lee & Tamborini, 2005), political parody videos on YouTube (Golan & Lim, 2016), and social media content (Dohle & Bernhard, 2014). To our knowledge, no study has yet examined the effect of third-person perception on support for restrictions of user comments. The current study would like to contribute further evidence:

H3: The stronger the users' third-person perception of a user comment, the stronger is the users' support for restrictions of the comment by authorities.

Studies considering the influence of a message's presumed influence on online corrective actions are limited in number (on offline corrective actions, e.g., Barnidge & Rojas, 2014; Wei, Chia, & Lo, 2011): Rojas (2010) indicates an effect of third-person perception on online information sharing and opinion expression. Golan and Lim (2016) show that third-person perception influences social media activism against ISIS's online recruitment propaganda. However, Lim (2017) cannot support this effect in a study on online advertising. Lim and Golan (2011) find that the perception of influence on others was a good

predictor in explaining participants' willingness to engage in political social media activism against political parody videos on YouTube. The current study aims at contributing further evidence. We postulate:

H4: The stronger the users' third-person perception on a user comment, the stronger is the users' intention to engage in corrective action against the comment.

A central theoretical explanation for the effect of third-person perception on rectifying behaviors is users' paternalistic orientation (Chia et al., 2004; McLeod et al., 2001; McLeod et al., 1997; Sun, Pan et al., 2008). Paternalism is a desire to keep others from hurting themselves, also by sometimes limiting their rights with a benign intention. This psychological phenomenon is believed to involve a feeling of self-superiority in that those manifesting paternalism believe to be less susceptible for negative influences and to know better what is good for others, whereas others are seen as more naïve, susceptible, and vulnerable. As such, paternalistic orientation is related to greater perceived influence on the more vulnerable others compared to oneself (McLeod et al., 1997). In a meta-analysis, the presumed vulnerability of the audience indeed turned out to be a significant moderator of third-person perception (Sun, Pan et al., 2008). Beyond that, the concept of paternalism also implies action taken on behalf of others, who are seen as in need of protection. Thus, paternalism is assumed to be directly related to support for censorship to finally protect others because of a concern for their well-being (McLeod et al., 2001).

H5: The higher a user's paternalistic tendency, the higher is their support for restrictions of an uncivil user comment by authorities.

To our knowledge, paternalism has not yet been related to the intention for corrective action to change or counterbalance media content. Nevertheless, it seems reasonable to assume that people with a greater tendency to protect others and a sense of superior ability to do so might more likely intent to engage against uncivil user comments beyond only

supporting restrictions by authorities since in social media they finally command means of intervention. Given the lack of existing research, we pose a research question:

RQ1: Does a user's paternalistic tendency influence their intention to engage in corrective action against a user comment?

Influence of Incivility on Third-Person Perception

Extensive research has identified numerous factors that determine this discrepancy between presumed influences on self and on others (for an overview, Sun, Pan et al., 2008). A key factor is that the severity of a message increases the third-person discrepancy (Gunther & Storey, 2003). This "negative influence corollary" is the result of a "self-serving bias" (Gunther & Mundy, 1993) or "self-enhancement bias" (Perloff, 2002). Individuals are eager to project a superior self-image and deny or downplay their susceptibility to messages with *undesirable* influences. The perceptual gap is assumed to diminish when the message is perceived as more socially desirable (e.g., Gunther & Mundy, 1993; Lim, 2017; Sun, Pan et al., 2008; Zhong, 2009; evidence from experimental tests with manipulations of the undesirable character of the message is provided by Eveland & McLeod, 1999; Scherr & Müller, 2017[1]).

Ample literature suggests that uncivil messages are perceived as undesirable. They threaten the face of the interaction partners, make them feel disrespected and restricted, and cause negative feelings (see above, e.g., Chen & Lu, 2017; Coe et al., 2014; Gervais, 2017; Muddiman, 2017). Therefore, we suggest that uncivil user comments cause a stronger third-person perception than civil comments. However, not all studies find empirical support for such relationship (Lambe & McLeod, 2005). About user comments, Chen and Ng (2016) argue that uncivil comments would lose persuasive power *because* they threaten the readers' sense of face and are perceived as unjustified and outside societal norms. Jenkins and Dragojevic (2011) relate face theory to resistance to persuasion. They indicate that people

perceive demeaning and controlling messages as face threats and may react to these messages by derogating the source and the particular message. Chen and Ng (2017) indeed show that uncivil comments are presumed to have a weaker effect on others' opinion than civil comments (the authors only test the presumed effect on others, not on self). This suggests that incivility could even lead to a weaker third-person perception.

To summarize, research in the area of uncivil media messages provides a heterogeneous picture: While third-person research generally assumes a stronger third-person discrepancy with more socially undesirable messages, face theory suggests a decreased presumed effect of more uncivil messages. Hence, we ask a research question:

RQ2: Does an uncivil user comment lead to a different third-person perception compared to a civil comment?

Influence of Commitment to Freedom of Expression

Paternalism seems to be a plausible psychological explanation of why people wish to rectify media content when they perceive it as uncivil and influential. However, focusing only on a mechanism that explains the motivation to counter a message would disregard competing mechanisms that motivate acceptance of messages – even despite these messages being uncivil and influential. Thus, in the pursuit to develop an integrated model support for restrictive actions and intention to corrective actions, we consider people's commitment to freedom of expression as additional, but countering influential factor.

Free speech is a central value in Western societies allowing individuals to express their mind without fear. It is protected to foster the free development of the individual and serve the collective good in a democratic society. Most Western constitutions grant freedom of expression even to extreme messages and messages perceived undesirable by some. However, the kind of privileges, that freedom of expression enjoys, varies across Western countries (for a comparison of US and Germany, Nieuwenhuis, 2002). In Germany, the

country of the present study, and in other European countries, anti-democratic utterances do not automatically enjoy the same protection as any other political utterances. Grounded in the experiences with the terror of the German National Socialist regime in the 1930s and 1940s, the German notion of free speech includes that "Democracy must be protected against anti-democratic groups, and freedoms must be protected against dictatorial movements" (Nieuwenhuis, 2002, p. 200). By this understanding, the German government passed the Network Enforcement Act in July 2017. The law governs conditions under which social media platforms that exceed a certain number of users will have to take down hate speech (Leisegang, 2017). Different from that, the US understanding of freedom of expression follows the notion of the free market place of ideas which demands that people must be able to consider every opinion. State interference in comment sections runs counter to this understanding.

Beyond the constitutional interpretation of freedom of expression and its institutional guarantee, there is the appreciation of freedom of expression on the individual level. Some people are more able or willing to consider the value of free speech as a general good to individuals and society. People who generally value freedom of expression more, are more likely to grant this freedom to extreme statements and communicators, and show less approval of restrictions. Differences between individuals in their commitment to freedom of expression arise not only from different cultural backgrounds but also from demographic, psychological, and sociopolitical differences (Bahry, Boaz, & Gordon, 1997; Gibson, 2006; Hurwitz & Mondak, 2002; Lambe, 2004; Paek et al., 2008; McLeod, Sotirovic, Voakes, Guo, & Huang, 1998).

When considering interventions against user comments in social media, the commitment to freedom of expression should play a role of inhibiting all too comprehensive limitations of user-generated content. Appreciation of the principle of free speech should

motivate weighting the downsides of media restrictions. Even with the more restrictive German notion of freedom of expression, any restriction to free speech is viewed with suspicion (Nieuwenhuis, 2002). Indeed, people who value freedom of expression higher, are more inclined to concede it even to undesirable or extreme expressions (Davis, 1990; Lambe, 2004; Marcus, Sullivan, Theiss-Morse & Wood, 1995; Paek et al., 2008; Price, Tewksbury, & Huang, 1998; Zhao & Cai, 2008). Thus, we postulate:

H6: The higher a user's general commitment to freedom of expression, the lower is the user's support for restrictions by authorities of an uncivil user comment.

Lim (2017) puts forward that "many socially harmful messages, even if they are extremely harmful and undesirable […], receive some constitutional protection in the United States, which limits the efficacy of legal regulation" (p. 976). The users can thus complement legal regulation with corrective actions – but their stance toward the protection of the constitution should not be overlooked. However, literature has not yet touched upon the question of whether the commitment to freedom of expression also affects individuals' willingness to engage in corrective actions against messages they perceive as undesirable or influential. Moreover, the motivation of German users to complement legal regulation has not been examined. Given the lack of empirical research, we pose the question

RQ3: Does a user's general commitment to freedom of expression influence their intention to engage in corrective action against an uncivil user comment?

Theoretical Model

Figure 1 summarizes the assumed relationships in a conceptual model. Beyond the determinants under investigation, further factors might plausibly influence support for restrictive and intention for corrective actions against a user comment. To control for these influences, the model incorporates two additional constructs: In general, people are more likely to tolerate content that is in line with their values and supports their attitudes (Huh et

al., 2006; Price et al., 1998). People who depreciate a group, person, or idea that is under attack in a user comment will perceive the user comments as less uncivil and less likely favor regulation of such comments (Kalch & Naab, 2017). Thus, we will control for the participants' attitudes towards the subject of the comment (i.e., refugees, see below). Additionally, the model includes participatory Facebook use. Readers who generally engage more often actively in social media, are more experienced and will also engage more likely in specific situations (Naab et al., 2018).

[Insert Figure 1 here]

Method

Design and procedure

We carried out a between-subjects design, varying incivility of a user comment (civil vs. uncivil). The participants filled a web-based survey. Integrated into the survey, they read a mock news post reporting on the number of refugees in Germany. The post was said to originate from the well-known and highly reputable German news outlet *SpiegelOnline* and said to be posted on the outlet's Facebook page. The post and the comments section were presented in the original layout of the *SpiegelOnline* Facebook page, yet did not contain personalized or interactive elements. We choose the topic of refugees because it had dominated news cycles at the time of the study. As a consequence of ongoing violent conflicts in Arab and African countries, a large number of refugees sought asylum in Germany provoking controversial debates about handling this immigration increase and thus also eliciting adversarial online comments. The presented news post had received three user comments. Two comments consisted of neutral statements and were identical in both conditions. We manipulated the third comment in two different versions (Table 1): The civil comment took a stand against receiving refugees in Germany and was formulated forthrightly and explicitly negative against refugees. In the uncivil condition, we added elements of

incivility (similar Gervais, 2017). The uncivil comment contained insulting, derogatory, and vulgar expressions, exaggeration, and included a highly discriminatory perspective against refugees (following Coe et al., 2014; Papacharissi, 2004).[2] The news post and the comments were based on actual messages on social network sites. After the experiment, the participants were thoroughly debriefed.

[Insert Table 1 here]

Participants

We recruited participants via mailing lists, emails, and postings in a wide variety of Facebook groups to enable a broad sample. Two hundred twenty-two participants were randomly assigned to one of the two conditions and answered the online questionnaire. To strengthen the external validity of the results, our analyses refer to the 213 participants who indicated at least rare usage of user comments (34.3% male; age: M = 28.30, SD = 11.95; 7.5% of the sample hold no school qualification or a secondary school certificate at best, 46.9% hold a university (of applied science) entrance certificate, 44.1% finished a university (of applied sciences) program, 1.4% did not answer the question).

Measures

Unless stated otherwise, participants answered on a scale from 1 = *fully disagree* to 7 = *fully agree*.

Support for restrictive actions by authorities. To measure support for restrictive actions by Facebook moderators and by the law enforcement agency, participants indicated their reaction in case they would find this comment on Facebook on a scale from 1 = *not at all* to 7 = *fully*. The items concerning actions by Facebook moderators were: Moderators should delete this comment, block the account of the comment author, check future comments of this author prior to publication, check future comments of this author after publication, use automatic computerized control of comments, call the law enforcement

agency into action. The mean index of the items had acceptable reliability (α = .86; *M* = 3.70; *SD* = 1.67). The items concerning actions by the law enforcement agency were: The law enforcement agency should prosecute the comment author, force Facebook to block the account of the comment author, force Facebook to delete the comment, check future comments of this author prior to publication, check future comments of this author after publication. The mean index had acceptable reliability (α = .88; *M* = 3.06; *SD* = 1.70). In the later analysis, the support for restrictive actions by authorities was modeled as a latent construct influencing two indicators, namely actions by Facebook moderators and by the law enforcement agency. It should be noted that not all of the restrictive actions are necessarily carried out. However, since the study attempts to investigate the perspective of the users, it includes potential instruments users might support independent of their application.

Intention to engage in corrective actions. Participants indicated on a scale from 1 = *not at all* to 7 = *fully* how likely they are to rebuke the comment author in a private message, rebuke the comment author in a public reply, report the comment with the flagging button, report it to the law enforcement agency, make other users aware of the comment, so that they react, report it in a private message to the page moderators. We computed a mean index (α = .80; *M* = 2.33; *SD* = 1.30).

Third-person perception. Participants rated the presumed effect of the comment on their knowledge, attitudes, and behaviors toward the German refugee policy and refugees with five items (α = .87; *M* = 1.68; *SD* = 1.09). Additionally, they rated the presumed effect on others on equivalent items (α = .91; *M* = 4.02; *SD* = 1.46). The respondents presumed the comment would have a significantly greater effect on others than on themselves, $t(212) = 20.330$, $p < .0001$. The third-person perception was computed as the difference between the presumed effect on others and on self, with higher values indicating a greater presumed effect on others and a theoretical range between -6 and +6 (*M* = 2.34; *SD* = 1.68).

General commitment to freedom of expression was measured with six items (Miklikowska, 2011; Rojas et al., 1996, e.g., "No matter how controversial the idea is, an individual should be able to express it publicly"). A mean index was computed (α = .73; *M* = 4.74; *SD* = 1.12).[3]

Paternalism was measured with three items (McLeod et al., 2001, e.g., "Sometimes it is necessary to protect people from harming themselves"; α = .65; *M* = 5.63; *SD* = .88).[4]

Attitudes toward refugees was measured with six items (adapted from Zick & Preuß, 2016, e.g., "The high number of refugees endangers the future of Germany"). High values indicate negative attitudes (α = .86; *M* = 3.08; *SD* = 1.33).

Participatory Facebook use was measured with three items. Participants indicated how often they a) comment on articles, b) reply to comments of other users, and c) post reactions to comments of other users (1 = *never* to 8 = *very often*). The items were aggregated to a mean index (α = .81; *M* = 2.47; *SD* = 1.59).

Manipulation check. Eight items on *incivility* asked participants how much they perceived the comment to be insulting, impolite, vulgar, derogatory, to infringe personality rights, infringe human rights, discriminate, and exaggerate (α = .92; *M* = 5.32; *SD* = 1.58).[5]

Results

The model was assessed with structural equation modeling explaining the support for restrictive actions by authorities and the intention to engage in corrective actions. Maximum likelihood (ML) estimates for the model were calculated using AMOS (Figure 2). The model fits the data well (*n* = 213; *p* = .143; CMIN(12) = 17.188; RMSEA = 0.045; PCLOSE = 0.519; CFI = 0.989; SRMR = 0.0413). Bootstrap standard errors and bias-corrected 95% confidence intervals were generated based on 2,000 bootstrap samples. For simplification purposes, error terms, and covariances between exogenous variables were omitted from the figure. The model allows for all plausible correlations between exogenous variables, but no

correlations with the manipulated variable. Due to their skewed distribution, the variables paternalism and participatory Facebook use were z-transformed before inclusion in the structural equation model.

[Insert Figure 2 here]

The model shows significant, direct effects of the manipulated variable on support for restrictive actions (β = .479; *p* = .001) as well as on intention to engage in corrective actions (β = .301; *p* = .001). Thus, H1 and H2 receive support. Albeit, the level of incivility does not influence third-person perception (β = .000 *p* = .996; RQ2). Third-person perception influences neither support for restrictive actions (β = .002; *p* = .990) nor intention for corrective actions (β = -.078; *p* = .214). H3 and H4 are not supported. Supporting H5, we find a significant, direct, positive effect of paternalism on the support for restrictive actions (ß = .150; *p* = .018). We find no direct effect of paternalism on the intention to engage in corrective actions (β = .080; *p* = .214; RQ1). Paternalism does not show indirect effects. Commitment to freedom of expression has a significant, negative effect on support for restrictive actions by authorities (β = -.263; *p* = .001). Supporting H6, the more important people consider free speech, the less they perceive intervention by Facebook moderators and law enforcement as adequate. However, appreciation of freedom of expression does not influence the intention to engage in corrective actions (β = .051; *p* = .364; RQ3).

Considering the control variables, the more people have negative attitudes toward refugees above average, the less they consider restrictive actions of authorities relevant (β = -.283; *p* = .001) and the less they intend to engage in corrective actions themselves (β = -.338; *p* = .001). People who generally engage more actively in comment spaces do not support restrictive actions more or less than less engaged users (β = .066; *p* = .382). However, people who generally engage more actively in comment spaces more likely intend to engage in corrective actions against the specific comment (β = .414; *p* = .001).

Discussion

The paper provided a unique contribution to the scholarship and the practice of social media regulation because it expanded the view to corrective actions as a potential consequence beyond support for restrictive actions by authorities. It addressed the fact that in today's media environment, users are less dependent on the regulation practices of authorities. The empowered public can directly respond to what it may regard as severely uncivil or harmfully influential social media content.

The study had an interest in understanding the role of incivility of media content and third-person perception in inducing restrictive and corrective actions. As expected, the incivility level of the user comment had a direct effect on support for restrictions and intention to engage in corrective actions. Additionally, comment readers with more negative attitudes towards the group attacked in the user comments deemed restrictive actions less necessary and intended fewer corrective actions themselves. This is in line with research stating that perceiving media content as hostile and biased against one's view leads to restrictive and corrective actions (Feldman et al., 2017; Hwang et al., 2008; Rojas, 2010; Wei et al., 2011). However, contrary to expectations, support for restrictions by authorities did not increase with increasing self-other discrepancy of the influence of the user comments. The model test did also not yield a significant path from third-person perception to intention to correction action. It seemed irrelevant for support for media regulations whether others are presumably strongly influenced. Thus, the study supported findings that show that the evaluation of a message per se has a stronger or exclusive influence than the presumption of effect (Bernhard & Dohle, 2014). This finding seems plausible when realizing that post hoc corrective and restrictive actions are sanctions for communication misbehavior and not necessarily measures to prevent an effect on the recipients (such as a priori censorship of media content).

The results further expand the literature on corrective action. Differentiating between restrictive and corrective actions turned out particularly fruitful since the determinants of both constructs partly differed. The findings suggested that support for authoritative regulation is a function of people's tendency for paternalism. People who assumed more than the average of the participants to know best what is good for others and were more willing to limit others' rights to prevent them from harmful consequences, were more inclined to ask for institutional restrictions of user comments. In contrast, paternalism did not lead users to a higher intention to perform corrective actions themselves. Future studies might want to investigate whether user regulation is perceived less effective and appropriate to protect vulnerable readers. This might also help news organizations to understand if they should actively engage in regulation in comment threads to signal to their readers that they provide a safe space for discussions.

People who commit to freedom of expression opposed restrictions by authorities. However, their appreciation of freedom of expression did not determine their stand toward corrective actions by the user. It seems that a commitment to freedom of expression in the abstracts makes people suspicious about institutionalized intervention, while it does not influence trust in the deliberate nature of the unregulated debate. It is a difficult task for platform providers and news organizations to find a balance between providing safe spaces for vulnerable groups but at the same time not limiting free debate about controversial topics. Surely, this balance is specific for each community and needs careful observation of the comment streams of each news outlet.

Participatory social media use was introduced as a control variable in the model. It turned out to be the most important influential factor in explaining the intention to engage in corrective actions. Thus, experience in producing user-generated content and giving feedback to others motivated people to act against problematic content. This result is in line with findings of online bystander research indicating that knowledge about intervention strategies

increases corrective action (Naab et al., 2018). It also suggests that platform providers and news organization who wish to involve users in the regulation of problematic comments might for one thing explicitly encourage regular users to apply their experience to intervene in comment threads. For another thing, they might want to point out the need for intervention and provide precise correction strategies to less experienced users. Finally, it remains the task of future research to explore the extent to which the now habitual handling of social media by numerous users has led to the establishment of behavioral norms about users' baseline engagement within these media.

The present study served as a pioneer effort to concurrently test a comprehensive model of determinants of restrictive and corrective actions, including both paternalism and appreciation of freedom of expression. It seemed that the two constructs are opposing determinants of support for restrictive actions by authorities. Users with a disposition for paternalism appreciate institutionalized media control while a commitment to freedom of expression makes people cautious about extensive intrusions of authorities.

Consistent with an extensive body of knowledge regarding self-other perceptual gaps, the results of our study supported the perceptual component of the third-person effect. Individuals were more likely to perceive others as more susceptible to uncivil user comments than themselves. Given the scientific evidence that uncivil comments indeed can have detrimental effects on individuals' attitudes and online and offline behavior (Anderson et al., 2014; Hwang & Kim, 2016; Thorson et al., 2010), it is noteworthy that people recognized the potential influence of uncivil user comments. However, they also showed the self-serving bias found with many other mediated messages, neglecting or at least understating effects of incivility on themselves. Albeit, the results did not support the negative influence corollary — the magnitude of the third-person perception presumably being affected by the undesirability of the message. While this relation is a robust finding in many studies, other

studies have also found that the severity of a message does not necessarily correlate with the attributed influence (Lambe & McLeod, 2005; McLeod et al., 2001; Paek et al., 2008). Additionally – and this lends substance to the recent findings – the effect of message undesirability on third-person perception has rarely been tested experimentally. The current study used a between-subject experimental design that manipulated the level of incivility of a user comment, keeping other features of the message and its context constant. The current finding substantiates claims by Chen and Ng (2016) who suppose that uncivil comments may lead to resistance to persuasion and lower presumed vulnerability of others because they might be perceived as transgressing societal norms of tolerable interaction style and thus less effective.

Limitations and Future Research

Several factors undermine the validity of our results: First, the study used a mock Facebook page and investigated the intention to engage in corrective actions in a hypothetical situation. This procedure aimed at increasing the internal validity of the results. However, it limits ecological validity because it does not measure user behavior in a real-world scenario of an interactive and personalized social network site.

Although the study adapted scales for commitment to freedom of expression and paternalism from literature that have proved to be successful, the low internal consistencies of the scales in the present study suggest to replicate the findings on the influences of these constructs with improved measures.

The user comments referred to a highly controversial debate in Germany about online hate speech and refugee discrimination. Surely further studies with other thematic contexts are necessary for generalizability. Additionally, different from the manipulated user comments, comments can also include less blatant forms of discrimination, justification

strategies, or references to presumed opinion climate. These may affect readers' perception of severity, the presumption of influence, and resistance to persuasion.

The design focused on comments on a Facebook news page. Such discussions are vitally relevant to news engagement and public discourse. As a complement, future studies might want to test whether the present results are generalizable to further media content and to other platforms and providers who offer different instruments of intervening with corrective actions and moderate to varying degrees.

Despite the broad recruitment, the sample was not representative of the German Facebook users. Primarily, the respondents have a higher level of education than the actual Facebook user community. Although interest in online news is greater among well-educated users (Hölig & Hasebrink, 2015), it would be rash to generalize the findings to all users of Facebook news pages. At least, research indicates that the likelihood to intervene against inappropriate behavior is less dependent on socio-demographics (Fischer et al., 2011). While the experimental setting and the convenient sample of our study are not about “making generalizations, but testing them” (Mook, 1983, 380; Meltzer, Naab & Daschmann, 2012), future studies need to replicate the findings with representative samples.

Conclusion

Sadly, uncivil user comments accompany many online discussions in social media. While social media platform providers and state authorities intervene with technical means and legal sanctions, the users, too, can engage against deviant comments. In social media, for the first time, users are equipped with considerable instruments to correct perceived harmful messages by writing their messages, liking or disliking the content of others or connect to further users or moderators. Their actions complement institutionalized regulatory actions. Therefore, the present study did not only contribute knowledge to the scholarship on

corrective action but also the practice of social media regulation. It provided insights into the users' assessment of who should engage in problematic comment threads.

References

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect": Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Bahry, D., Boaz, C., & Gordon, S. B. (1997). Tolerance, transition, and support for civil liberties in Russia. *Comparative Political Studies*, *30*(4), 484–510. https://doi.org/10.1177/0010414097030004004

Barnidge, M., & 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. https://doi.org/10.1093/ijpor/edt032

Bernhard, U., & Dohle, M. (2014). Do even journalists support media restrictions? Presumed political media influences and the consequences. *Journalism and Mass Communication Quarterly*, *91*(2), 250–271. https://doi.org/10.1177/1077699014527456

Bernhard, U., & Dohle, M. (2015). Corrective or confirmative actions? Political online participation as a consequence of presumed media influences in election campaigns. *Journal of Information Technology & Politics*, *12*(3), 285–302. https://doi.org/10.1080/19331681.2015.1048918

Brett, J. M., Olekalns, M., Friedman, R., Goates, N., Anderson, C., & Lisco, C. C. (2007). Sticks and stones: Language, face, and online dispute resolution. *Academy of Management Journal*, *50*(1), 85–99. https://doi.org/10.5465/amj.2007.24161853

Brosius, H.-B., & Engel, D. (1996). The causes of the third-person effect: Unrealistic optimism, impersonal impact, or generalized negative attitudes towards media influence? *International Journal of Public Opinion Research*, *8*(2), 142–162. https://doi.org/10.1093/ijpor/8.2.142

Brown, P., & Levinson, S. (1987). *Politeness. Some universals in language usage*. Cambridge, UK: Cambridge University Press.

Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. Cham: Springer International Publishing.

Chen, G. M., & Ng, Y. M. M. (2016). Third-person perception of online comments: Civil ones persuade you more than me. *Computers in Human Behavior*, *55*, 736–742. https://doi.org/10.1016/j.chb.2015.10.014

Chen, G. M., & Ng, Y. M. M. (2017). Nasty online comments anger you more than me, but nice ones make me as happy as you. *Computers in Human Behavior*, *71*, 181–188. https://doi.org/10.1016/j.chb.2017.02.010

Chia, S. C., Lu, K.-H., & McLeod, D. M. (2004). Sex, lies, and video compact disc. *Communication Research*, *31*(1), 109–130. https://doi.org/10.1177/0093650203260204

Chung, M., Munno, G. J., & Moritz, B. (2015). Triggering participation: Exploring the effects of third-person and hostile media perceptions on online participation. *Computers in Human Behavior*, *53*, 452–461. https://doi.org/10.1016/j.chb.2015.06.037

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

Dahlberg, L. (2011). Re-constructing digital democracy: An outline of four 'positions'. *New Media & Society*, *13*(6), 855–872. https://doi.org/10.1177/1461444810389569

Davis, J. A. (1990). Attitudes towards free speech in six countries in the mid-1980: Australia, Austria, Great Britain, Italy, the United States and West Germany. *European Sociological Review*, *6*(1), 1–14.

Davis, P., Morrison, G., Johnson, M. A., & Ross, F. (2002). Body image, race, and fashion models social distance and social identification in third-person effects. *Communication Research, 29*(2), 270-294. https://doi.org/10.1177/0093650202029003003

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly, 47*(1), 1. https://doi.org/10.1086/268763

Dohle, M., & Bernhard, U. (2014). Presumed online media influence and support for censorship: Results from a survey among German parliamentarians. *International Journal of Public Opinion Research*, *26*(2), 256–268. https://doi.org/10.1093/ijpor/edt027

Driscoll, P. D., & Salwen, M. B. (1997). Self-perceived knowledge of the O. J. Simpson trial: Third-person perception and perceptions of guilt. *Journalism and Mass Communication Quarterly*, *74*(3), 541–556. https://doi.org/10.1177/107769909707400308

Duggan, M., & Smith, A. (2016). The political environment on social media. Retrieved from http://www.pewinternet.org/2016/10/25/the-political-environment-on-social-media/

Eveland, W. P., & McLeod, D. M. (1999). The effect of social desirability on perceived media impact: Implications of third-person perceptions. *International Journal of Public Opinion Research*, *11*(4), 315–333. https://doi.org/10.1093/ijpor/11.4.315

Feldman, L., Hart, P. S., Leiserowitz, A., Maibach, E., & Roser-Renouf, C. (2017). Do hostile media perceptions lead to action? The role of hostile media perceptions, political efficacy, and ideology in predicting climate change activism. *Communication Research, 44*(8), 1099–1124. https://doi.org/10.1177/0093650214565914

Feng, G. C., & Guo, S. Z. (2012). Support for censorship: A multilevel meta-analysis of the third-person effect. *Communication Reports*, *25*(1), 40–50. https://doi.org/10.1080/08934215.2012.661019

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., . . . Kainbacher, M. (2011). The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*(4), 517–537. https://doi.org/10.1037/a0023304

Gervais, B. T. (2017). More than mimicry? The role of anger in uncivil reactions to elite political incivility. International Journal of Public Opinion Research, 29(3), 384-405. https://doi.org/10.1093/ijpor/edw010

Gibson, J. L. (2006). Enigmas of intolerance: Fifty years after Stouffer's communism, conformity, and civil liberties. *Perspectives on Politics*, *4*(1), 21–34. https://doi.org/10.1017/S153759270606004X

Goffman, E. (1967). *Interactional ritual: Essays on face-to-face behaviour*. Harmondsworth, Middlesex, England: Penguin.

Golan, G. J., & Lim, J. S. (2016). Third-person effect of ISIS's recruitment propaganda: Online political self-efficacy and social media activism. *International Journal of Communication*, *10*, 4681–4701.

Graham, T., & Wright, S. (2015). A tale of two stories from "below the line". *The International Journal of Press/Politics*, *20*(3), 317–338. https://doi.org/10.1177/1940161215581926

Grice, H. P. (1969). Utterer's meaning and intentions. *The Philosophical Review*, *78*(2), 147–177.

Gunther, A. C., & Mundy, P. (1993). Biased optimism and the third-person effect. *Journalism Quarterly*, *70*(1), 58–67. https://doi.org/10.1177/107769909307000107

Gunther, A. C., & Storey, J. D. (2003). The influence of presumed influence. *Journal of Communication*, *53*(2), 199–215. https://doi.org/10.1111/j.1460-2466.2003.tb02586.x

Hoffner, C., & Buchanan, M. (2002). Parents' responses to television violence: The third-person perception, parental mediation, and support for censorship. *Media Psychology*, *4*(3), 231–252. https://doi.org/10.1207/S1532785XMEP0403_02

Hölig, S. & Hasebrink, U. (2015). *Reuters Digital News Survey 2015. Ergebnisse für Deutschland [Reuters Digital News Survey 2015: Results for Germany]*. Retrieved from http://www.hans-bredow-institut.de/webfm_send/1095

Hoorens, V., & Ruiter, S. (1996). The optimal impact phenomenon: Beyond the third person effect. *European Journal of Social Psychology*, *26*(4), 599–610. https://doi.org/10.1002/(sici)1099-0992(199607)26:4<599::aid-ejsp773>3.3.co;2-z

Houston, J. B., Hansen, G. J., & Nisbett, G. S. (2011). Influence of user comments on perceptions of media bias and third-person effect in online news. *Electronic News*, *5*(2), 79–92. https://doi.org/10.1177/1931243111407618

Huh, J., Delorme, D. E., & Reid, L. N. (2006). Perceived third-person effects and consumer attitudes on preventing and banning DTC advertising. *Journal of Consumer Affairs*, *40*(1), 90–116. https://doi.org/10.1111/j.1745-6606.2006.00047.x

Hurwitz, J., & Mondak, J. J. (2002). Democratic principles, discrimination and political intolerance. *British Journal of Political Science*, *32*(1), 93–118. https://doi.org/10.1017/S0007123402000042

Hwang, H., & Kim, Y. (2016). Influence of discussion incivility on deliberation: An examination of the mediating role of moral indignation. *Communication Research, published online before print.* https://doi.org/10.1177/0093650215616861

Hwang, H., Pan, Z., & Sun, Y. (2008). Influence of hostile media perception on willingness to engage in discursive activities: An examination of mediating role of media indignation. *Media Psychology*, *11*(1), 76–97. https://doi.org/10.1080/15213260701813454

Jenkins, M., & Dragojevic, M. (2011). Explaining the process of resistance to persuasion: A politeness theory-based approach. *Communication Research, 40*(4), 559.590. https://doi.org/10.1177/0093650211420136

Jensen, J. D., & Hurley, R. J. (2005). Third-person effects and the environment: social distance, social desirability, and presumed behavior. *Journal of Communication, 55*(2), 242–256. https://doi.org/10.1111/j.1460-2466.2005.tb02670.x

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

Lambe, J. L. (2004). Who wants to censor pornography and hate speech? *Mass Communication and Society, 7*(3), 279–299. https://doi.org/10.1207/s15327825mcs0703_2

Lambe, J. L., & McLeod, D. M. (2005). Understanding third-person perception processes: Predicting perceived impact on self and others for multiple expressive contexts. *Journal of Communication, 55*(2), 277–291. https://doi.org/10.1111/j.1460-2466.2005.tb02672.x

Lee, B., & Tamborini, R. (2005). Third-person effect and internet pornography: The influence of collectivism and internet self-efficacy. *Journal of Communication, 55*(2), 292–310. https://doi.org/10.1111/j.1460-2466.2005.tb02673.x

Leisegang, D. (2017). No country for free speech?: An old libel law and a new one aimed at social media are two threats to free expression in Germany. *Index on Censorship, 46*(2), 76–78. http://doi.org/10.1177/0306422017716038

Lim, J. S. (2017). The third-person effect of online advertising of cosmetic surgery: A path model for predicting restrictive versus corrective actions. *Journalism and Mass Communication Quarterly, 94*(4), 972–993. https://doi.org/10.1177/1077699016687722

Lim, J. S., & Golan, G. J. (2011). Social media activism in response to the influence of political parody videos on YouTube. *Communication Research, 38*(5), 710–727. https://doi.org/10.1177/0093650211405649

Marcus, G. E., Sullivan, J. L., Theiss-Morse, E. & Wood, S. L. (1995). *With malice toward some. How people make civil liberties judgments.* Cambridge, UK: Cambridge University Press.

McLeod, D. M., Detenber, B. H., & Eveland, W. P. (2001). Behind the third-person effect: Differentiating perceptual processes for self and other. *Journal of Communication, 51*(4), 678–695. https://doi.org/10.1111/j.1460-2466.2001.tb02902.x

McLeod, D. M., Eveland, W. P., & Nathanson, A. I. (1997). Support for censorship of violent and misogynic rap lyrics: An analysis of the third-person effect. *Communication Research, 24*(2), 153–174. https://doi.org/10.1177/009365097024002003

McLeod, J. M., Sotirovic, M., Voakes, P. S., Guo, Z., & Huang, K.-Y. (1998). A model of public support for First Amendment rights. *Communication Law & Policy, 3*(4), 479–514. https://doi.org/10.1080/10811689809368662

Meltzer, C., Naab, T. & Daschmann, G. (2012). All student samples differ: Participant selection in communication science. *Communication Methods and Measures, 6*(4), 1-12. https://doi.org/10.1080/19312458.2012.732625

Miklikowska, M. (2011). *Democracy begins at home: Parenting, empathy, and adolescents' support for democratic values*. Vasa: KTMP.

Mook, D. G. (1983). In defense of external validity. *American Psychologist, 38*(4), 379-387. http://dx.doi.org/10.1037/0003-066X.38.4.379

Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication, 11*, 3182–3202.

Naab, T. K., Kalch, A., & Meitz, T. G. (2018). Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on bystander behavior. *New Media & Society, 20*(2), 777–795. https://doi.org/10.1177/1461444816670923

Neuwirth, K., & Frederick, E. (2002). Extending the framework of third-, first-, and second-person effects. *Mass Communication and Society, 5*(2), 113–140. https://doi.org/10.1207/S15327825MCS0502_2

Nieuwenhuis, A. (2000). Freedom of speech: USA vs Germany and Europe. *Netherlands Quarterly of Human Rights*, *18*(2), 195–214. https://doi.org/10.1177/092405190001800203

Oetzel, J. G., Ting-Toomey, S., Yokochi, Y., Masumoto, T., & Takai, J. (2000). A typology of facework behaviors in conflicts with best friends and relative strangers. *Communication Quarterly*, *48*(4), 397–419. https://doi.org/10.1080/01463370009385606

Paek, H.-J., Lambe, J. L., & McLeod, D. M. (2008). Antecedents to support for content restrictions. *Journalism and Mass Communication Quarterly*, *85*(2), 273–290. https://doi.org/10.1177/107769900808500204

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

Paul, B., Salwen, M. B., & Dupagne, M. (2000). The third-person effect: A meta-analysis of the perceptual hypothesis. *Mass Communication and Society*, *3*(1), 57–85. https://doi.org/10.1207/S15327825MCS0301_04

Perloff, R. M. (2002). The third-person effect. In J. Bryant (Ed.), *Media effects. Advances in theory and research* (pp. 489–506). Mahwah, NJ: Erlbaum.

Price, V., Tewksbury, D., & Huang, L.-N. (1998). Third-person effects on publication of a Holocaust-denial advertisement. *Journal of Communication*, *48*(2), 3–26. https://doi.org/10.1111/j.1460-2466.1998.tb02745.x

Rojas, H. (2010). "Corrective" actions in the public sphere: How perceptions of media and media effects shape political behaviors. *International Journal of Public Opinion Research*, *22*(3), 343–363. https://doi.org/10.1093/ijpor/edq018

Rojas, H., Shah, D. V., & Faber, R. J. (1996). For the good of others: Censorship and the third-person effect. *International Journal of Public Opinion Research*, *8*(2), 163–186. https://doi.org/10.1093/ijpor/8.2.163

Rosenthal, S., Detenber, B. H., & Rojas, H. (2018). Efficacy beliefs in third-person effects. *Communication Research*, *45*(4), 554–576. https://doi.org/10.1177/0093650215570657

Rösner, L., Winter, S., & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, *58*, 461–470. https://doi.org/10.1016/j.chb.2016.01.022

Scherr, S., & Müller, P. (2017). How perceived persuasive intent and reactance contribute to third-person perceptions: Evidence from two experiments. *Mass Communication and Society*, *20*(3), 315–335. https://doi.org/10.1080/15205436.2016.1250911

Shah, D. V., Faber, R. J., & Youn, S. (1999). Susceptibility and severity. Perceptual dimensions underlying the third-person effect. *Communication Research*, *26*(2), 240–267. https://doi.org/10.1177/009365099026002006

Stroud, N. J., Scacco, J. M., Muddiman, A., & Curry, A. L. (2015). Changing deliberative norms on news organizations' Facebook sites. *Journal of Computer-Mediated Communication*, *20*(2), 188–203. https://doi.org/10.1111/jcc4.12104

Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility? *Communication Monographs*, *83*(4), 535–556. https://doi.org/10.1080/03637751.2016.120120

Sun, Y., Pan, Z., & Shen, L. (2008). Understanding the third-person perception: Evidence from a meta-analysis. *Journal of Communication*, *58*(2), 280–300. https://doi.org/10.1111/j.1460-2466.2008.00385.x

Sun, Y., Shen, L., & Pan, Z. (2008). On the behavioral component of the third-person effect. *Communication Research*, *35*(2), 257–278. https://doi.org/10.1177/0093650207313167

Thorson, K., Vraga, E., & Ekdale, B. (2010). Credibility in context: How uncivil online commentary affects news credibility. *Mass Communication and Society*, *13*(3), 289–313. https://doi.org/10.1080/15205430903225571

Thurman, N., Cornia, A., & Kunert, J. (2016). *Journalists in the UK.* Available from https://reutersinstitute.politics.ox.ac.ukTsfati, Y., Ribak, R., & Cohen, J. (2005). Rebelde way in Israel: Parental perceptions of television influence and monitoring of children's social and media activities. *Mass Communication and Society*, *8*(1), 3–22. https://doi.org/10.1207/s15327825mcs0801_2

Tsfati, Y., Ribak, R. & Cohen, J. (2005). Rebelde way in Israel. Parental perceptions of television influence and monitoring of children's social and media activities. *Mass Communication and Society, 8*(1), 3–22. https://doi.org/10.1207/s15327825mcs0801_2

Wei, R., Chia, S. C., & Lo, V.-H. (2011). Third-person effect and hostile media perception influences on voter attitudes toward polls in the 2008 U.S. presidential election. *International Journal of Public Opinion Research*, *23*(2), 169–190. https://doi.org/10.1093/ijpor/edq044

Wei, R., & Lo, V.-h. (2007). The third-person effects of political attack ads in the 2004 U.S. presidential election. *Media Psychology*, *9*(2), 367–388. https://doi.org/10.1080/15213260701291338

Xu, J., & Gonzenbach, W. J. (2008). Does a perceptual discrepancy lead to action? A meta-analysis of the behavioral component of the third-person effect. *International Journal of Public Opinion Research*, *20*(3), 375–385. https://doi.org/10.1093/ijpor/edn031

Zhao, X., & Cai, X. (2008). From self-enhancement to supporting censorship: The third-person effect process in the case of internet pornography. *Mass Communication and Society*, *11*(4), 437–462. https://doi.org/10.1080/15205430802071258

Zhong, Z.-J. (2009). Third-person perceptions and online games: A comparison of perceived antisocial and prosocial game effects. *Journal of Computer-Mediated Communication*, *14*(2), 286–306. https://doi.org/10.1111/j.1083-6101.2009.01441.x

Zick, A., & Preuß, M. (2016). Einstellungen zur Integration in der Bevölkerung [Attitudes on integration in the population]. Retrieved from https://www.stiftung-mercator.de/media/downloads/3_Publikationen/Zugleich_Kurzbericht_2016.PDF

Endnotes

[1] Many further studies rather assume the undesirability of a message (e.g., Golan & Lim, 2016) than measure or even manipulate it. Some studies measure perceived social desirability (e.g., Jensen & Hurley, 2005; Lim, 2017) or use various topics with natural variations of negative severity (e.g., Gunther & Mundy, 1993; Hoorens & Ruiter, 1996). Others compare the perception of people with opposing attitudes towards the topic indicating reversed perceptions of the desirability of the same message (Driscoll & Salwen, 1997). Further studies vary the framing of the message as being desirable or undesirable (Brosius & Engel, 1996). Some regard a single message and measure the presumed effects on various, more or less desirable outcomes (Davis, Morrison, Johnson, & Ross, 2002).

[2] Although formulated very directly and impolite, the civil stimulus comment does not include such markers of incivility that separate uncivil from "merely" impolite speech like discrimination and undemocratic content. Therefore, we decided to name the conditions "civil" and "uncivil" (for a differentiation between impoliteness and incivility see Papacharissi, 2004).

[3] Internal consistence of the scale could not be improved by dropping items. A principal component factor analysis was conducted on the six items with oblique rotation (direct oblimin). The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .77, and all KMO values for individual items were at least 0.70. One factor had an eigenvalue over Kaiser's criterion of 1 and explained 53.70% of the variance.

[4] Internal consistence of the scale could not be improved by dropping items. A principal component factor analysis was conducted on the three items with oblique rotation (direct oblimin). The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .65, and all KMO values for individual items were at least 0.64. One factor had an eigenvalue over Kaiser's criterion of 1 and explained 59.06% of the variance.

[5] A *t*-test showed that the uncivil comment was perceived as more uncivil (n = 100, M = 6.26, SD = 1.17) than the civil (but still negative against refugees) comment (n = 113; M = 4.49, SD = 1.42), $t(209.937) = -10.01$, $p < .0001$. This indicates that the manipulation of incivility was successful.

To control for the validity of the setting, we measured *investigative quality* (M = 4.55, SD = 1.26), *comprehensibility* (M = 5.66, SD = 1.33), and *credibility of the news post* (M = 5.05, SD = 1.38) as well as *realism* (M = 5.72, SD = 1.87) and *authenticity of the comments* (M = 5.72; SD = 1.50). Differences between experimental conditions were neither visible for perceived investigative quality, $t(211) = -.12$, $p = .907$, comprehensibility, $t(211) = 1.46$, $p = .145$, and credibility of the news post, $t(211) = -0.87$, $p = .388$, nor for the perceived realism, $t(211) = -0.29$, $p = .771$, and authenticity of the comment section, $t(211) = -0.17$, $p = .867$.

Table

Table 1

Stimulus comments

Civil comment	**Uncivil comment**
Just so that you guys understand me, I'm not a Nazi, but I stand by my opinion. **Not all of them that come are bad**, but enough is enough! A lot of **people** come here just to get our money and other things and to take advantage of us. **They** come and want work and money and **complain** afterwards that it's all not enough. And us Germans are **left out in the cold**! All you hear is, you need to adapt, accept, tolerate, … **They** should not be here in the first place, let alone demand for anything! Integration **is failing** because they have no respect for us and our culture. They all need to **be deported**!!	Just so that you guys understand me, I'm not a Nazi, but I stand by my opinion. We've got enough of **these parasites** here. Enough is enough! **These bootlickers** come here just to get our money and other things and to take advantage of us. **Just like a plague of locusts**, more and more people are coming and want work and money and **moan** afterwards that it's all not enough. And us Germans **get the shit end of the stick**! All you hear is, you need to adapt, accept, tolerate, … **The dirty criminal asylum pack** should not be here in the first place, let alone demand for anything! **Fuck** integration! They have no respect for us and our culture. They all need to **be disposed of as genetic hazardous waste**!!

Note:

The study was conducted in Germany. The original German comments were translated for publication.

Main differences between the comments boldfaced.

Figures

Figure 1

Theoretical model of the influences on support for restrictive actions by authorities and intention to engage in corrective actions

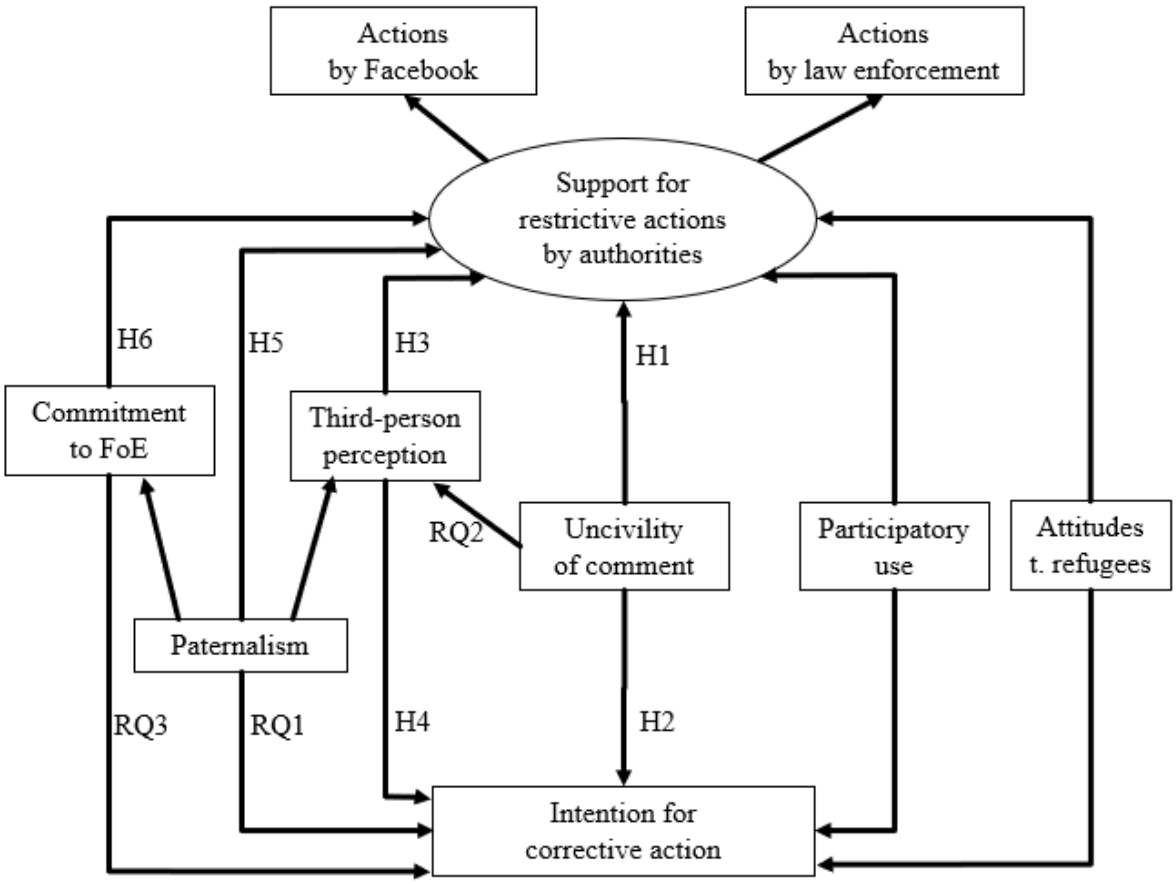
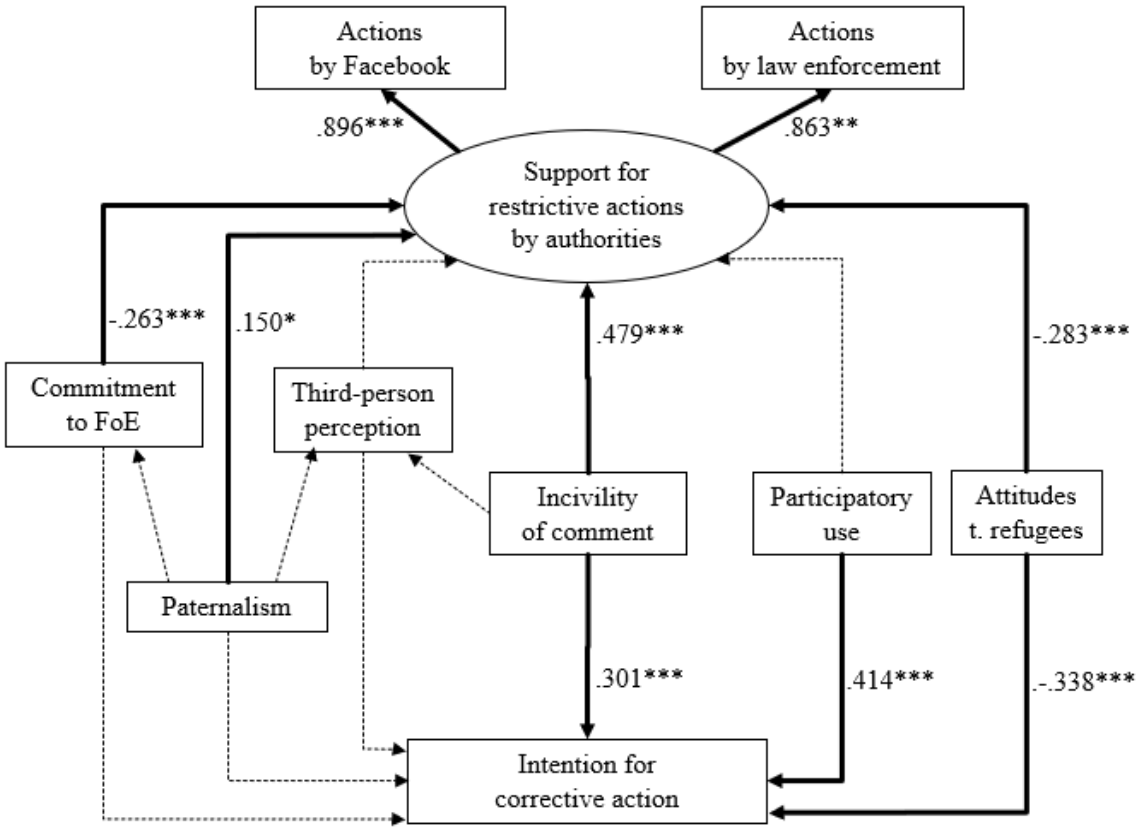


Figure 2

Structural equation model of the influences on support for restrictive actions by authorities and intention to engage in corrective actions



n = 213; *p* = .143; CMIN(12) = 17.188; RMSEA = 0.045; PCLOSE = 0.519; CFI = 0.989; SRMR = 0.0413

*$p \leq .05$; **$p \leq .01$; ***$p \leq .001$

Dashed lines mark non-significant paths ($p > .05$).