Replying, disliking, flagging: how users engage with uncivil and impolite comments on news sites

Anja Kalch, Teresa K. Naab

Angaben zur Veröffentlichung / Publication details:

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

SCM
Studies in Communication and Media

FULL PAPER

Replying, disliking, flagging: How users engage with uncivil and impolite comments on news sites

Antworten, Bewerten, Melden: Wie NutzerInnen gegen unzivile und unhöfliche Kommentare auf Nachrichtenwebseiten intervenieren

Anja Kalch & Teresa K. Naab

Anja Kalch, Department of Media, Knowledge, and Communication, Augsburg University, Universitätsstraße 10, 86159 Augsburg, Germany; Contact: anja.kalch(at)phil.uni-augsburg.de

Teresa K. Naab, Department, Department of Media, Knowledge, and Communication, Augsburg University, Universitätsstraße 10, 86159 Augsburg, Germany; Contact: teresa.naab(at)phil.uni-augsburg.de

FULL PAPER

Replying, disliking, flagging: How users engage with uncivil and impolite comments on news sites

Antworten, Bewerten, Melden: Wie NutzerInnen gegen unzivile und unhöfliche Kommentare auf Nachrichtenwebseiten intervenieren

Anja Kalch & Teresa K. Naab

Abstract: Uncivil and impolite user comments diminish the quality of deliberative discussions in the comment sections of news sites. This study investigates how users engage against such comments. Replying, disliking, and flagging are distinguished as different options of user engagement that vary in their functionality for discussion sections. To investigate the effects of incivility and impoliteness on user engagement, we conducted a 2 (civil vs. uncivil user comment) x 2 (polite vs. impolite user comment) online experiment. Results show that users engaged against impolite comments that attacked Muslims by flagging or by writing a reply against the language style or a reply against the expressed opinion. Incivility did not influence user engagement. The effects are moderated by Islamophobic attitudes. The results give detailed insights into readers' behavior in discussion threads and have consequences for professional moderation.

Keywords: User comments, user engagement, incivility, impoliteness, experiment, news sites, Muslims

Zusammenfassung: Unzivile und unhöfliche Nutzerkommentare schwächen den deliberativen Diskurs in den Kommentarbereichen von Nachrichtenwebseiten. Diese Studie untersucht, wie NutzerInnen gegen solche Kommentare intervenieren. Das Schreiben von Antwortkommentaren, Bewerten und Melden werden als Mittel des Nutzerengagements differenziert, die sich in ihren Auswirkungen auf den Diskussionsverlauf unterscheiden. Um die Einflüsse von Unzivilität und Unhöflichkeit auf das Nutzerengagement zu untersuchen, haben wir ein 2 (ziviler vs. unziviler Kommentar) x 2 (höflicher vs. unhöflicher Kommentar) Online-Experiment durchgeführt. Die Ergebnisse zeigen, dass NutzerInnen unhöfliche Kommentare, die Muslime angreifen, sanktionieren, indem sie diese an Moderatoren melden oder einen kritisierenden Antwortkommentar schreiben. Islamophobe Einstellungen der NutzerInnen moderieren den Einfluss. Unzivilität hat dagegen keinen Einfluss auf das Nutzerverhalten. Die Ergebnisse geben detaillierten Einblick in das Nutzerverhalten in Online-Diskussionen und haben Konsequenzen für professionelle Moderatoren.

Schlagwörter: Nutzerkommentare, Online Engagement, Unzivilität, Unhöflichkeit, Experiment, Nachrichtenwebseiten, Muslime

1. Introduction

Online discussions in comment sections on news sites have the potential to increase deliberative discourse by connecting disperse audiences. From a normative claim, online discussions on news sites should represent diverse viewpoints, be respectful, polite, rational, coherent, sincere, and comprehensible (e.g., Graham & Witschge, 2003; Stromer-Galley, 2007). Such discourse is assumed to influence knowledge, opinions, decision-making processes, democratic attitudes, and civic engagement in a positive manner (Mutz, 2008). However, comment sections are often criticized for not fulfilling the claims for a deliberative discourse (e.g., Coe, Kenski, & Rains, 2014; Reich, 2011). Therefore, some news outlets have completely shut down their comment sections, while others limit the comment function to specific topics (e.g., Jensen, 2016). Still, many platforms offer opportunities for user participation, and professional moderators supervise discussion threads prior to or post-publication (Goodman, 2013; Noci, Domingo, Masip, Micó, & Ruiz, 2010; Reich, 2011).

However, professional moderation faces challenges, and a closer look at the potential contributions of user engagement in the regulation process seems worthwhile (Naab, 2012, 2016a): The sheer number of comments that need the consideration of professional moderators calls for an effort requiring both staff and financial outlay. In many countries, providers of comment sections are even legally required to check and eventually delete comments if they have been made aware of the potentially illegal content. Users can contribute to the moderation process by expressing their opinions and assisting professional moderators in regulatory decisions. User opinions can also legitimate decisions by moderators (Reich, 2011). Engagement by a large number of users reduces the risk of limiting certain viewpoints because control is dispersed to many judges (Crawford & Gillespie, 2016; Noci et al., 2010). Concertive control among the users can increase cohesion and participation and in return increase user awareness about the rules of the comment section (Gibbs, Kim, & Ki, 2016). Additionally, user engagement seems imperative in online spaces that are not checked by professional moderators, for example, when news items are discussed outside the official comment space of news outlets.

Given these possibilities of user engagement, it seems worthwhile to draw scholarly attention to the engagement of users against deviant comments of other lay authors. However, the variety of more or less inappropriate content in user comments is broad. While impolite comments, including insults or abusive language, may be easily detected, uncivil content threatening democratic values may be even more disruptive to deliberative discourse but also harder to detect. It might be even more challenging to identify and engage against "impeccable incivility," which comes in well-mannered, polite language (Papacharissi, 2004). This should be relevant in particular for attacks against people or groups that already suffer from prejudice and stereotypes so that attacks are at risk of being easily tolerated. In addition to the variety of inappropriate comments, users also can and have to choose between various options of engagement against such com-

ments. These options differ in their discursive contribution to the discussion and in their consequences for follow-up professional moderation.

This paper will particularly consider the nature of incivility and impoliteness of online comments and test the influences and interactions of these characteristics on user engagement. The paper examines different types of user engagement in comment sections, namely, replying, pushing dislike buttons, and flagging comments. Given the assumed influence of prior attitudes, we discuss Islamophobia as a moderator for effects of comments attacking Muslims. From the findings, we draw conclusions on user sanctions in online interactions and on moderation practices in comment sections.

2. Uncivil and impolite user comments

Many comment sections contain positive discussions of high value for readers and platform providers (e.g., Graham & Wright, 2015). Still, although inappropriate comments are not in the majority in most comment sections, they are a reason for complaints by readers and journalists and a challenge for moderators (Coe et al., 2014; Santana, 2014; Stroud, Scacco, Muddiman, & Curry, 2015). To differentiate in more detail between inappropriate comments and to aim at a more fine-grained understanding of user engagement against such comments, we will draw a distinction between uncivil and impolite comments following the work by Papacharissi (2004). She proposes that incivility manifests itself in "behaviors that threaten democracy, deny people their personal freedoms, and stereotype social groups" (p. 267). Uncivil comments may, for example, contain racist content, sexist content, content otherwise refusing equal rights to all, or attacking people for belonging to certain ethnic or social groups. These attacks can be directed against members participating in the discussion as well as against other absent people such as subjects in media coverage (Papacharissi, 2004). From a normative position based on democratic pluralism, such uncivil comments are categorically intolerable. Empirical literature on the effects of user comments supports the assumption that uncivil comments may have negative effects on readers' attitudes and contribute to spreading undemocratic opinions (Ballantine, Lin, & Veer, 2015; von Sikorski, 2016), increase attitude certainty, and decrease open-mindedness, but also motivate willingness to political participation (Borah, 2014). Additionally, incivility in mediated messages other than user comments (e.g., political campaign statements) has been shown to decrease perceived importance and informational value of messages, but also to increase political interest and political participation (Brooks & Geer, 2007).

While incivility is defined based on undemocratic content and is independent of language, impoliteness is shown in informal speech style, non-compliance to any sort of etiquette, and noncooperation in conversation (Papacharissi, 2004; for detailed elaborations on politeness, see also Brown & Levinson, 1987). Extreme examples of impoliteness are name-calling, casting aspersions, pejorative speech, and vulgarity (Papacharissi, 2004). Impoliteness in comment sections can be directed against other discussants as well as against individuals or groups not directly participating in a discussion, such as politicians and media actors (Pa-

pacharissi, 2004). In terms of effects, impolite comments can lead to reduced open-mindedness (Hwang & Kim, 2016), polarization of attitudes (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014), and more negative perceptions of the communicator (Ng & Detenber, 2005; on the effects of impoliteness in media content other than comments e.g., Mutz & Reeves, 2005).

Politeness generally is context-dependent; a particular style may be perceived desirable by one group or under some circumstances but inappropriate by others. Group norms about an adequate communication style develop as discussions proceed (Sukumaran, Vezich, McHugh, & Nass, 2011; Stroud et al., 2015). Still extreme forms of impoliteness violate the usage policies of most comment sections, and many scholars consider them to hamper deliberative discourse (mostly based on the conception of Habermas, e.g., 1989). However, some authors emphasize that even with aggressiveness, humiliations, and insults online discussions could be attentive to opposing arguments and exercise free speech. Too much desire for a well-mannered discourse and a rejection of passionate arguments could downplay the value of dissent and lead to self-censorship and the suppression of opposing views (Benson, 1996; Oetzel et al., 2003). In a similar vein, Rossini (2017) points out that rude remarks, bad manners, and a disrespectful tone do not necessarily prevent a discussion from being democratically relevant. Still, uncivil expressions, even if formulated in a polite tone, are always dysfunctional for democracy.

Both characteristics in comments – incivility and impoliteness – violate the usage policies of most comment spaces. Both can be assumed to elicit negative sanctions by engaged users, too. In line with that, Coe and colleagues (2014) show that users of comment sections evaluate uncivil and impolite comments more negatively than appropriate comments. Naab, Kalch, and Meitz (2016) find more flagging of deviant comments. The findings by Stryker and colleagues (2016) also support the idea that people perceive both incivility and impoliteness as problematic.

It is unclear so far whether impoliteness and incivility could have different effects on user engagement in accordance with the conceptual differences outlined above. Regarding the perception of incivility and impoliteness, Stryker and colleagues (2016) show that, for example, racial and sexist slurs – apparently uncivil expressions by our definition – are perceived as more deviant than insults, name-calling, and other impolite expressions. This would suggest that users engage more likely in a consequent way against uncivil comments than impolite ones. However, users might more easily identify extreme forms of impoliteness than uncivil content. Most users might perceive extreme forms of impoliteness as inappropriate in comment sections directed to a general audience and thus more likely sanction impolite comments. Classifying comments as uncivil requires a reflection on democratic and social values (Papacharissi, 2004) and might also depend on the opinion toward the object of attack held by the particular reader.

Evaluating comments as uncivil and threatening to democratic pluralism might be further complicated when such content is expressed firmly and is well-elaborated. Uncivil discussions may be more easily perceived as problematic when they also contain rude language, which is an apparent violation of discussion rules (Papacharissi, 2004). Research about political discussions supports this assumption, showing that incivility in combination with impoliteness negatively affects

credibility perceptions and attitudes (Brooks & Geer, 2007; Thorson, Vraga, & Ekdale, 2010). However, research has not specifically investigated interaction effects between impoliteness and incivility on user engagement yet. Following from the above elaboration we hypothesize:

H1: Users will more likely engage against uncivil than against civil user comments.

H2: Users will more likely engage against impolite than against polite user comments.

RQ1: Is user engagement influenced by the interaction of impoliteness and incivility?

3. User engagement against uncivil and impolite comments: Replying, disliking, and flagging

Users have several options to engage with other users' comments on news sites and pages of news organizations in social networks. So far, empirical research has barely addressed how flagging, button evaluations, and replies are related to each other. It is likely that these options are not used independently of one another, given their different meaning, the required effort, and their functionality for professional moderation. Most probably, users decide for the option that is most suitable for their needs. Up to now, there is only limited evidence how different types of inappropriate comments motivate replying, disliking, and flagging.

Users can **reply to the comments** of others and engage in a discursive exchange of opinions. However, regarding the likelihood of replying against inappropriate comments, empirical results are heterogeneous. Ziegele (2016) shows that users engage in negative responses mainly when comments express views opposing their own opinion, but also to indicate a lack of rationality and comprehensibility or to complain about inappropriate redundancy and aggression. Singer (2009) finds that readers indeed quickly engage against inappropriate comments. In contrast, Coe and colleagues (2014) show that readers barely reply to critical comments. These differences may be based on variations in usage policies of platforms and their expectations for users: Some platform providers advise readers not to engage in inappropriate comments to prevent flame wars and not to fuel trolls, while others point to the necessity to side against undemocratic attitudes by counterarguing (Kühl, 2015). Besides varying platform policies, the users might have reasons to reply or not to reply to impolite or uncivil comments, too. Users may, for example, have different expectations about how effective a reply to impolite or uncivil comments is. Thus, we ask the research question:

RQ2: How is writing a reply related to uncivil and impolite comments?

By **flagging comments**, users can report violations of the usage policy to professional moderators, who may eliminate or change problematic content (Crawford & Gillespie, 2016). Regulations by professional moderators have direct and immediate consequences for the sanctioned author. Thus, flagging can be more effective than replying and button evaluations (Naab et al., 2016). The flagging alert is

particularly important to providers who do not systematically monitor all comments. Providers can increase the likelihood of flagging inappropriate comments through providing information that briefs the users about the meaning and usage of the flagging button (Naab et al., 2016). However, a flag has a limited contribution to deliberative discourse since the authors of flagged comments are usually not informed about being flagged and have no opportunity to object or justify themselves. The meaning of a flag is even more questionable as several reasons besides violations of usage policies can lead to flagging, including pranks or bullying. At the same time, not flagging inappropriate content (and the same may hold true for not using evaluation buttons and not replying) is not necessarily a signal of agreement but may indicate ambivalence toward the content, inertia, a lack of knowledge, or a lack of perceived self-efficacy (Crawford & Gillespie, 2016). Naab and colleagues (2016) show that flagging is more likely for clearly deviant comments that attack users directly than for less deviant comments.

RQ3: How is flagging related to uncivil and impolite comments?

One of the most common types of user engagement is *liking* an object. Some news outlets also provide dislike buttons (Engelmann & Marzinkowski, 2017) or similar evaluation buttons (e.g., Stroud, Muddiman, & Scacco, 2016). On Facebook, where readers also can comment on news articles, users can even choose from a variety of so-called reactions with positive and negative valence. In contrast to writing a reply, an evaluation button is a less effortful type of user engagement but also allows users to express one's approval or disapproval in public (Hölig & Hasebrink, 2015; Sarapin & Morris, 2014). Pushing buttons indicates "that one is reading comments, even though not participating in the actual debate" (Eranti & Lonkila, 2015, p. 10). The opportunity to evaluate content with the push of a button can increase people's willingness to express their opinion, especially when they disagree with the aggregated rating of others (Hong & Park, 2011). However, evaluation buttons only indicate a summary statement but do not give room for reasoning and justifications, and are thus a less discursive option of sanctioning than writing a reply. Since social button counters present the aggregated ratings of all users, the evaluated author cannot negotiate with an individual evaluator. Regarding the usage of likes and dislikes, the content of a posting is the most important reason for liking an object, while uncivil comments receive more negative button evaluations (Coe et al., 2014).

In contrast to flagging that is supposed to report violations of the usage policy, disliking reflects the standpoint of the users and their involvement with the topic. Experiences of Disqus, a provider of discussion platforms also hosting comment sections of online newspapers, support the assumption that people are aware of the difference between disliking and flagging. Disqus has noticed that many users flag to express disagreement with a stated opinion when no dislike button is available. When an additional button has been introduced in a comment section to evaluate comments negatively, the amount of flagging decreased considerably (Goodman, 2013). This indicates that users conceptually differentiate between disliking and flagging. However, the question is how such a conception of disliking manifests itself for impolite and uncivil comments.

RQ4: How is pushing the dislike button related to uncivil and impolite comments?

4. The moderating influence of topic-related attitudes on user engagement

Not only the content of a comment is likely to impact user engagement but also attitudes and values reflecting the users' standpoint toward the topic in question. Based on a content analysis of Usenet posts, Papacharissi (2004) shows that uncivil content is often based on strong prior attitudes and values. Likewise, readers will interpret such content against the backdrop of their attitudes. The perception that an uncivil user comment opposes democratic values should be stronger for individuals with positive attitudes toward the object under attack of the comment (Borah, 2014). The same should hold true for the perception of impoliteness when it is directed against an object that is valued by the individual reader. In contrast, uncivil but polite comments may particularly appear likeminded to people holding negative attitudes toward the offended subject (Gervais, 2015).

The assumed influence of attitudes on perceptions of incivility and impoliteness is in line with research showing that the perceived hostility of media content depends on the personal position of the readers (Borah, 2014). Also, bystander research has shown that people more likely help victims to whom they have a closer relationship (Levine & Crowther, 2008), have more frequent contact, and toward whom they hold less prejudice (Abbott & Cameron, 2014). Considering user comments, positive attitudes toward a social group attacked by a user comment have been shown to increase flagging behavior of inappropriate comments (Naab et al., 2016; Stroud et al., 2016).

The influence of attitudes and values of the users is even more relevant for attacks against people or groups that already suffer from widespread prejudice and stereotypes because attacks might be less scrutinized or even tolerated. Among others, this may be the case for Muslims in Western democracies. Over the last few years, discrimination of Muslims in the U.S. and Western European countries has increased (Ogan, Willnat, Pennington, & Bashir, 2014). This "anti-Muslim bias" (Bansak, Hainmueller, & Hangartner, 2016, p. 1) is not restricted to right-wing voters but is also common in other parts of the population. Prejudice and critical views toward Muslims manifests itself in Islamophobic attitudes (Imhoff & Recker, 2012). Given that Islamophobia influences the standpoint of users toward Muslims and their democratic rights, it is likely to affect user engagement against uncivil and impolite comments directed against Muslims. Thus, we assume:

H3: Islamophobia moderates the effects of uncivil and impolite user comments directed against Muslims on user engagement. In contrast to people with positive attitudes toward Muslims, people with negative attitudes toward Muslims will less likely condemn uncivil and impolite user comments.

5. Method

5.1 Design and participants

A 2 x 2 between-subjects design, varying the civility (civil vs. uncivil) and politeness (polite vs. impolite) of a user comment was carried out. Participants were randomly assigned to one of the four conditions of a news site and answered a self-administered follow-up online questionnaire.

We conducted the study in Germany. A convenience sample was used. Students of a communication class, who were extensively trained in social science methods, recruited participants via mailing lists, e-mails, and postings in social networking sites. Participants received no compensation for participation. Overall, 160 people participated.

We excluded three participants (1.8%) who indicated they never read user comments on news sites or social network sites. Doing this should strengthen the external validity of the results because these participants might have limited knowledge and never be in a situation to engage against inappropriate comments. The analysis refers to 157 participants (M_{age} = 26.06, SD = 10.44, 55 males (35.0%), 93 females (59.3%), nine people did not indicate gender (5.7%)). Most of the participants have a higher level of education (86.7% had the general qualification for university entrance).

5.2 Stimuli

A fully functional comment section on a news site that enabled user engagement was programmed (see Appendix 1 and Appendix 2). The participants read an article of the politics section since politics is known to motivate many users to write comments but also to attract worse user comments (Coe et al., 2014; Goodman, 2013). The news article described the imprisonment and torture of the Saudi Arabic blogger Raif Badawi in the context of democratic rights. The topic was discussed in the German media in December 2015 and January 2016 when we conducted the study. The site and the comment section were designed in close resemblance to news sites of popular German online newspapers. The comment section included a note that it welcomed a vital, but well-mannered debate. It explained the available options of user engagement, informing users that they should flag comments that discriminate, provoke, or are perceived as problematic for other reasons and that professional moderators would take care of flagged comments. The introductory note also provided a link to a comprehensive netiquette. Additionally, it encouraged users to state their opinion either by using like and dislike buttons or by writing a reply.

In the comment section, three user comments were included. We used such a short thread to avoid confounding the variables as well as a superposition of effects. The thread length is equal to the beginning of a discussion. Two comments were not manipulated and consisted of neutral statements. We manipulated the last comment in four different versions. Following Papacharissi (2004), incivility and impoliteness can be directed against other discussants as well as against indi-

viduals or groups not directly participating in a discussion. We manipulated other-directed incivility and impoliteness attacking Arab Muslims in general for not intervening against human rights abuses. The impolite versions of the comment contained the use of insults, vulgarity, abusive language, and name-calling against Muslims. While the civil comment suggested legal prosecution of responsible politicians and inactive bystanders, the uncivil comment demanded the death penalty without referring to the right to a fair trial. Additionally, the uncivil comment stereotyped Muslims with reference to wearing yashmaks (face cover worn by Muslim women). The comments were formulated very extremely, because Naab and colleagues (2016) showed that flagging in comment sections is low and increases with the perceived deviance of a comment (see Appendix 3 for the comments).

Below each comment, three buttons were provided to enable liking (green thumbs up button), disliking (red thumbs down button), and flagging (red button labeled 'Report'). Additionally, a reply field was added to each comment, offering participants the opportunity to respond to a comment directly. At the end of the comment section, an open commentary field was included to enable participation in the discussion without replying to any of the prior comments.

Participants were asked to read the article and engage in the discussion as they would normally do. In the instruction, participants were informed that former participants of the study had written the comments included in the comment section and that comments they write would also be visible to later participants, similar to a real debate. This was to increase external validity by simulating the situation of interacting publicly. After the experiment, we fully debriefed the participants, revealing that the researchers had made up the comments and that their own comments would not be reused.

5.3 Measures

The site captured if the participants clicked (1) or did not click (0) the dislike button or the flagging button and if they wrote a reply to the manipulated comment or not. In total, 57 participants replied directly to the comment. Two coders coded all replies in two steps. First, whether a reply expressed a standpoint for or against the manipulated comment was coded. All replies argued against the comment. In a second step, they were coded regarding two categories: (1) whether they contained negative sanctions against the language of the comment, criticizing the insults and vulgarity of the comments and (2) whether they contained negative sanctions against the expressed opinion, criticizing discrimination and stereotyping as manipulated. Intercoder reliability was good ($Holsti_{style}$ = .95; $Holsti_{opinion}$ = .90).

Muslim-related attitudes were measured with a 15-item Islamophobia scale (Imhoff & Recker, 2012; e.g., "Islam is an archaic religion that is unable to adapt to the present," 1 = *strongly agree*, 7 = *strongly disagree*). High values represent a negative perception of Muslims (M = 3.94, SD = .81, α = .77).

In order to check if the manipulation was successful, four items on incivility asked participants how much they perceived the comment to infringe on personal rights, to infringe on human rights, to discriminate, and to be antidemocratic (1 = *fully disagree*, 7 = *fully agree*, M = 5.05, SD = 1.38, α = .75). Regarding impolite-

ness participants were asked to indicate how much the comment uses abusive language, vulgarism, casts aspersion, and is impolite (M = 5.55, SD = 1.65, α = .86). Additionally, believability of the article (M = 5.71, SD = 1.29), perceived authenticity of the comment section (M = 5.43, SD = 1.46), and realism of the comments (M = 5.00, SD = 1.59) were measured as controls for the setting (1 = *strongly disagree*, 7 = *strongly agree*).

6. Results

6.1 Manipulation check

A manipulation check was conducted to test whether the manipulations of incivility and impoliteness were successful. A t-test showed that the uncivil comments were perceived as more uncivil (M = 5.25, SD = 1.36) than the civil comments (M = 4.80, SD = 1.39), $t(155) = 2.06$, $p = .041$. The impolite comments were perceived as more impolite (M = 6.71, SD = .49) than the polite comments (M = 4.34, SD = 1.55), $t(90.62) = 12.82$, $p < .001$. No differences between experimental conditions were visible for believability of the article, $F(3, 153) = .11$, $p = .956$, perceived authenticity of the comment section, $F(3, 153) = .59$, $p = .621$, and realism of the comments, $F(3, 153) = .11$, $p = .957$.

6.2 Research questions and hypothesis testing

Overall, 21.7 percent (n = 34) of the participants did not engage with the comment, while most of the participants (78.3%, n = 123) showed at least one reaction toward the comment. Two-thirds of the people used one option of user engagement (64.2%, n = 79), one-third expressed their opinion by two options (35.8%, n = 44), and nobody used three options. Flagging (n = 46, 29.3%) and disliking (n = 50, 31.8%) were used more frequently than negative replies against the expressed opinion (n = 35, 22.3%) or against the language style (n = 21, 13.4%). We conducted correlation analyses (RQ1) to investigate the relationships between different options of user engagement. A significant negative correlation is visible between disliking and flagging (n = 157, r = -.200, p = .012). This indicates that nondiscursive types of user engagement are used rather exclusively. Negative replies criticizing language style or the expressed opinion are correlated with each other (n = 157, r = .194, p = .015), indicating that to some degree sanctions against the style and the stated opinion go hand in glove. No correlation is visible between negative replies and nondiscursive types of user engagement, indicating specific scopes of application for both categories of user engagement.

To refer to the theoretically discussed relationship and given the empirical correlations between the dependent variables (types of user engagement), a multivariate analysis of variance (MANOVA) was conducted. Impoliteness (polite vs. impolite) and incivility (civil vs. uncivil) were entered as independent variables and all four options of user engagement (flagging, disliking, negative reply against language style, negative reply against opinion) as dependent variables (H1, H2, and RQ1). The results show a multivariate effect of impoliteness on the user engage-

ment variables, $F(4, 150) = 8.70$, $p < .001$, $\eta^2_{part} = .19$, $\Lambda = .81$, but not of incivility or the interaction of impoliteness and incivility. Thus, H2 is supported but not H1. Regarding the univariate effects on different options of user engagement (RQ2–RQ4), impoliteness significantly effects flagging, $F(1, 153) = 15.63$, $p < .001$, $\eta^2_{part} = .09$. Impolite comments received significantly more flags ($M = 0.43$, $SD = 0.50$) than polite comments ($M = 0.16$, $SD = 0.37$). Similarly, impolite comments ($M = 0.20$, $SD = 0.40$) received significantly more negative replies against the language style, $F(1, 153) = 7.11$, $p = .008$, $\eta^2_{part} = .04$, than polite comments ($M = 0.06$, $SD = 0.25$). In contrast, polite comments ($M = 0.32$, $SD = 0.47$) received more negative replies against the expressed opinion, $F(1, 153) = 10.03$, $p = .002$, $\eta^2_{part} = .06$, than impolite comments ($M = 0.13$, $SD = 0.33$). Impoliteness does not show a significant effect on disliking. Again, no effects of incivility and no interaction effects on any of the options of user engagement are visible.

The assumed moderation effect of Islamophobia (H3) was tested using the PROCESS macro for SPSS, model 1. The experimental manipulation was entered as the multicategorical independent variable to analyze combined effects of impoliteness and incivility. Effect coding for the independent variable was used to draw comparisons of each comment with the grand group mean. Islamophobia was entered as moderator, and the variable was mean centered before the analysis. Bootstrap standard errors and bias-corrected 95% confidence intervals were generated based on 10,000 bootstrap samples. The analysis was conducted separately for flagging, disliking, replies against language style, and replies against the expressed opinion (Table 1).

Concerning flagging, both impolite comments (civil and uncivil) increased the likelihood of flagging, but only when people had positive or moderate attitudes toward Muslims and not when people had negative attitudes toward Muslims. In a similar vein, people with positive or moderate attitudes toward Muslims were less likely to flag the civil and polite comment. In general, the flagging likelihood of individuals holding negative attitudes toward Muslims is not influenced by the two comment characteristics.

Table 1. Indirect effects of comment characteristics on user engagement (flagging, disliking, reply against the expressed opinion, reply against the language style) at different levels of Islamophobia

	Low Islamophobia (*M* - 1 *SD*)		Moderate Islamophobia (*M*)		High Islamophobia (*M* + 1 *SD*)		Interaction term
Comment	Indirect effect (boot SE)	Boot 95 % CI	Indirect effect (boot SE)	Boot 95 % CI	Indirect effect (boot SE)	Boot 95 % CI	
flagging							
polite-civil	**-1.22 (0.53)**	**[-2.28, -0.17]**	**-1.44 (0.58)**	**[-2.58, -0.30]**	-1.65 (0.97)	[-3.55, 0.25]	*B* = -0.26 (*SE* = 0.64), *p* = .68
polite-uncivil	-0.54 (0.46)	[-1.44, 0.35]	-0.26 (0.36)	[-0.96, 0.44]	0.03 (0.53)	[-1.01, 1.07]	*B* = 0.35 (*SE* = 0.43), *p* = .40
impolite-civil	**0.83 (0.42)**	**[0.0005, 1.66]**	**0.86 (0.34)**	**[0.20, 1.52]**	0.89 (0.53)	[-0.15, 1.93]	*B* = 0.04 (*SE* = 0.43), *p* = .93
impolite-uncivil	**0.94 (0.47)**	**[0.009, 1.87]**	**0.83 (0.34)**	**[0.17, 1.49]**	0.73 (0.48)	[-0.20, 1.66]	*B* = -0.13 (*SE* = 0.42), *p* = .76
disliking							
polite-civil	-0.22 (0.43)	[-1.05, 0.62]	-0.16 (0.34)	[-0.83, 0.52]	-0.10 (0.53)	[-1.14, 0.95]	*B* = 0.07 (*SE* = 0.43), *p* = .86
polite-uncivil	**0.97 (0.40)**	**[0.19, 1.75]**	**0.64 (0.28)**	**[0.08, 1.20]**	0.31 (0.42)	[-0.51, 1.13]	*B* = -0.41 (*SE* = 0.36), *p* = .26
impolite-civil	-0.17 (0.44)	[-1.04, 0.69]	-0.21 (0.33)	[-0.84, 0.42]	-0.25 (0.50)	[-1.23, 0.74]	*B* = -0.04 (*SE* = 0.43), *p* = .92
impolite-uncivil	-0.58 (0.52)	[-1.60, 0.44]	-0.27 (0.33)	[-0.92, 0.37]	0.03 (0.40)	[-0.75, 0.81]	*B* = 0.38 (*SE* = 0.40), *p* = .35
reply against expressed opinion							
polite-civil	**0.86 (0.42)**	**[0.03, 1.69]**	**0.87 (0.35)**	**[0.18, 1.56]**	0.88 (0.58)	[-0.25, 2.01]	*B* = 0.01 (*SE* = 0.45), *p* = .97
polite-uncivil	-0.44 (0.48)	[-1.39, 0.51]	0.36 (0.34)	[-0.31, 1.02]	**1.15 (0.49)**	**[0.19, 2.11]**	***B* = 0.99 (*SE* = 0.43), *p* = .02**
impolite-civil	-0.27 (0.50)	[-1.26, 0.71]	-0.36 (0.41)	[-1.16, 0.43]	-0.45 (0.69)	[-1.80, 0.90]	*B* = -0.11 (*SE* = 0.55), *p* = .84
impolite-uncivil	-0.14 (0.58)	[-1.28, 1.00]	-0.86 (0.47)	[-1.78, 0.06]	**-1.59 (0.80)**	**[-3.15, -0.02]**	*B* = -0.89 (*SE* = 0.64), *p* = .16
reply against language style							
polite-civil	-15.85 (17630)	[-3457, 3454]	-16.12 (14863)	[-2914, 2912]	-16.40 (23533)	[-4614, 4611]	*B* = -0.34 (*SE* = 18032.45), *p* = 1.00
polite-uncivil	4.37 (5876.99)	[-1151, 1152]	4.83 (4954.49)	[-9705, 9715]	5.29 (7844.57)	[-1537, 1538]	*B* = 0.57 (*SE* = 6010.82), *p* = .99
impolite-civil	5.94 (5876.00)	[-1151, 1152]	5.77 (4954.49)	[-9704, 9716]	5.59 (7844.57)	[-1537, 1538]	*B* = -0.21 (*SE* = 6010.82), *p* = 1.00
impolite-uncivil	5.54 (5876.09)	[-1151, 1152]	5.53 (4954.49)	[-9705, 9716]	5.51 (7844.57)	[-1537, 1538]	*B* = -0.02 (*SE* = 6010.82), *p* = 1.00

Note. Values in boldface are significant at $p < .05$.

For disliking, only the polite but uncivil comment had an effect related to Islamophobic attitudes. The polite but uncivil comment received relatively more dislikes than the other three comments from people with positive or moderate attitudes toward Muslims. The two comment characteristics did not influence the likelihood of expressing a dislike of people holding negative attitudes toward Muslims.

Regarding replies against the expressed opinion, people with positive or moderate attitudes toward Muslims were more likely to write a reply against the expressed opinion of a polite and civil comment. Interestingly, people with negative attitudes toward Muslims were more likely to write replies against the opinion expressed in the polite but uncivil comment. But they were less likely to write a reply against the opinion expressed in the impolite and uncivil comment.

No conditional effects of Islamophobia are visible for replies against language style.

7. Discussion

Impoliteness and incivility of user comments are often mixed in the literature and are used interchangeably for several related but conceptually different types of inappropriate user comments. Following the differentiation by Papacharissi (2004), we elaborated on a distinction and showed that impoliteness and incivility are both problematic for deliberative discourse in comment sections. However, independently of the context, incivility is always dysfunctional to democracy, while impoliteness is a breach of etiquette. This conceptual differentiation is relevant in particular regarding the combination of incivility and impoliteness and in relation to different options of user engagement. Therefore, flagging, disliking, and writing a reply were differentiated as options of user engagement that vary in their contribution to democratic discourse and their expressiveness for professional moderation.

Overall, user engagement against the stimulus comments was mostly limited to one particular reaction, indicating that different options of user engagement are used in an exclusive and not in an accumulated way. This suggests that users do not want to overemphasize their standpoint by using several or all options simultaneously. User engagement with one exclusive option also implies that users attribute specific functionalities to replying, disliking, and flagging. Our results show that readers use flagging and disliking more often than negative replies against the stimulus comments. The smaller percentage of negative replies compared to the usage of social buttons is in line with existing literature (Hölig & Hasebrink, 2015; Ruiz et al., 2011). Writing a response requires more time and effort by the users than clicking a button.

Concerning the comment characteristics, only impoliteness of a comment influenced whether users engaged against a comment by flagging, by writing a reply that sanctioned its language style, or by writing a reply that countered the opinion expressed in the comment. Users engaged relatively strictly against impoliteness. In contrast, uncivil expressions were not detected. An inappropriate language usage seems to be more obvious for readers than incivility. For uncivil but polite comments this is in line with the notion of "impeccable incivility"

(Papacharissi, 2004), that is, when incivility is hidden in a well-mannered polite language. However, also in combination with impoliteness, incivility did not have a significant effect.

This indicates a challenging aspect for professional moderation: Neither when it comes in polite nor in impolite language do users identify incivility as problematic. This result is even more surprising since we used a relatively extreme example of incivility in the study. Efforts are necessary to increase the negotiation of shared values together with users, to make democratic values visible to the community, and to increase knowledge about different types of inappropriate and undemocratic arguments. The results of this study also question the reliability of user engagement. Given that users less likely condemn uncivil expressions, it seems that particular responsibility rests on professional moderators to be aware of less obvious threats against democracy and human rights. When users have little interest or low capabilities to identify problematic content that needs to be taken care of, professional moderators and platform providers need to find appropriate procedures to identify and delete such content before or shortly after publication without establishing too rigorous regulations limiting deliberative exchange. This is even more challenging given the liability of platform providers in many countries to correctly identify problematic content (Oozeer, 2014). While technical developments support professional moderators efficiently by detecting abusive language, hate speech, or further inappropriate content based on machine-learning algorithms, uncivil arguments are harder to detect using software algorithms given that they may be expressed in a well-mannered style.

Regarding the different types of user engagement, flagging and writing a reply were related to impoliteness, but disliking was not. In particular, flagging was used as the option to react against impolite comments. So flagging seems to be perceived the most effective way to sanction inappropriate comments. In contrast to disliking, flagging forcefully indicates a need for intervention to professional moderators, who can delete comments or even block users. Thus, flagging can be an immediate solution to the problem (Naab et al., 2016). Especially people with positive attitudes toward Muslims decided to intervene against the comments with offensive language against this social group by flagging. Presumably, people with positive attitudes toward Muslims perceived the comments as more deviant.

When we compare flagging to writing a reply comment, impolite comments increase the likelihood of a negative reply against the language style used in contrast to polite comments, but overall flags were used more often. Writing a reply as a direct reaction to an inappropriate comment requires comparably more time and effort but allows and requires readers to elaborate on the reasons for their critique in public.

While users react to impolite comments by writing response comments criticizing the language style, polite comments increase the likelihood that users will write a reply against the expressed opinion. People with a more positive attitude toward Muslims engaged more often in replies to polite and civil comments. People with negative attitudes toward Muslims responded more often to the polite and uncivil comment but less often to the impolite and uncivil comment. Thus, a polite language style seems to be an indicator for a valuable discussion that is

worthy of further participation. This is in line with content analytical results showing that user debates are most often rational debates (Graham & Wright, 2015; Rowe, 2014).

This study has some important limitations. The comments used in this study were formulated in a rather extreme way to motivate user engagement. Not only the uncivil but also the impolite comments expressed obvious depreciation of Muslims. While such extreme comments sadly fall into the scope of some recent online discussions, they address but one pole of the continuum from very uncivil and impolite to very respectful. This approach excludes much of the diversity of politeness and civility in user comment sections and limits the generalizability of the results. It may also lead to specific patterns of interaction between impoliteness and incivility. We did not find replies that supported the expressed opinion but all replies argued against the stimulus comment. This can surely be explained by the extreme comments used in the study that restrained the likelihood of expressing support. However, Graham and Wright (2015) also show that opposing claims are posted more often in a discussion than supporting claims. Furthermore, participants with negative attitudes toward Muslims did not reply positively (which of course might also be due to the laboratory situation). In general, the found usage patterns reflect the overly negative content of the comments and may be different for positive and high-quality comments or less extreme examples. Future research should go a step further and investigate whether lighter breaches of etiquette have a similar effect and result in negative sanctions, and in how far they can stimulate cognitive reflection and participation (Ikeda & Boase, 2011; Kim, 2016). However, researchers need to be aware that lighter forms of impoliteness might be perceived problematic in some contexts and by some participants and not by others (Stroud et al., 2015; Sukumaran et al., 2011).

Besides limitations in the stimulus material, further constraints of the study need attention. The replies of the participants may be more elaborated and less offensive than outside the laboratory. There is a clear need for field studies on user engagement. Still combining observation of social media behavior with survey data is challenging. Additionally, the generalizability of the findings needs to be confirmed for further topics. The topic used in the article may have limited active, discursive participation due to a low involvement of the readers with the situation of the Saudi Arabian blogger. Furthermore, the findings need to be confirmed for interventions to help other attacked social groups, whether these be further groups referred to in a news article or active users in the comment section.

The setting of the comment section used in this study allowed users to like and to dislike comments, to flag comments, to reply to previous comments, and to write new comments. This is a fairly comprehensive choice of options for user engagement. Often fewer options are available to users; for example, in some forums no dislike button is integrated (Engelmann & Marzinkowski, 2017). A different setting would likely result in different engagement patterns as would changes in the labels of the buttons. Future research thus is in need to investigate the effects of the setting on user behavior (for an example, see Stroud et al., 2016).

With regard to the sample, two limitations emerge. The sample of the study is rather young and well-educated. We focused on users who at least on occasion

read user comments on news sites or social network sites and excluded participants who indicated they never read or write user comments. Thus, we can draw conclusions for participants who under real-world conditions have the opportunity to engage against inappropriate comments. It is likely that this sampling strategy limited the distribution of sociodemographic variables in the sample because German users of comment sections are well-educated, too. Findings with regard to the age of German comment users are inconsistent (Hölig & Hasebrink, 2015; Springer, Engelmann, & Pfaffinger, 2015). This may point to a potential bias and limited generalizability of the results. However, bystander research has shown that sociodemographic variables have only little influence on helping attacked social groups (Fischer et al., 2011). In a similar vein, sociodemographics have been shown to explain comments behavior only to a small extent (Ziegele, Johnen, Bickler, Jakobs, Setzer, & Schnauber, 2013). Additionally, the distribution of Islamophobia found in the sample mirrors well the population regarding its attitudes toward Muslims (Zick, Küpper, & Hövermann, 2011). Beyond the generalizability of the existing results, it seems relevant to investigate further factors that could influence user engagement against inappropriate comments. The present study has focused on characteristics of the comments. However, demographics, personality traits (Downs & Cowan, 2012; Naab, 2016; Kenski, Coe, & Rains, 2017) as well as user experiences, perceived responsibility for the discussion, and commitment to the community (Baumeister, Chesner, Senders, Tice, 1988; Naab et al., 2016) have also been shown to affect deviance perception and intervention behavior.

8. Conclusion

The relevance of user-generated comments on professional media coverage for the perception of media content, the image of news providers, and for online and offline civic participation underline the importance of considering closely how to guide the quality of discussion threads. The present study provided insight that user engagement can complement professional moderation but is limited with regard to uncivil content. Users intervene against impolite comments. This is an important step toward ensuring deliberative discussions. However, incivility is not condemned that clearly. Strategies need to be discussed for ways to best moderate threats against democracy and human rights that are formulated in a well-mannered way.

References

Abbott, N., & Cameron, L. (2014). What makes a young assertive bystander? The effect of intergroup contact, empathy, cultural openness, and in-group bias on assertive bystander intervention intentions. *Journal of Social Issues*, *70*(1), 167–182. doi:10.1111/josi.12053

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. doi:10.1111/jcc4.12009

Ballantine, P. W., Lin, Y., & Veer, E. (2015). The influence of user comments on perceptions of Facebook relationship status updates. *Computers in Human Behavior*, *49*, 50–55. doi:10.1016/j.chb.2015.02.055

Bansak, K., Hainmueller, J., & Hangartner, D. (2016). How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers. *Science, 354*(6309), 217–222. doi:10.1126/science.aag2147

Baumeister, R. F., Chesner, S. P., Senders, P. S., & Tice, D. M. (1988). Who's in charge here? *Personality and Social Psychology Bulletin, 14*(1), 17–22. doi:10.1177/0146167288141002

Benson, T. W. (1996). Rhetoric, civility, and community: Political debate on computer bulletin boards. *Communication Quarterly*, *44*(3), 359–378. doi:10.1080/01463379609370023

Borah, P. (2014). Does it matter where you read the news story? Interaction of incivility and news frames in the political blogosphere. *Communication Research*, *41*(6), 809–827. doi:10.1177/0093650212449353

Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science, 51*(1), 1–16. doi:10.1111/j.1540-5907.2007.00233.x

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

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. doi:10.1111/jcom.12104

Crawford, K., & Gillespie, T. (2016). What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society*, *18*(3), 410–428. doi:10.1177/1461444814543163

Downs, D. M., & Cowan, G. (2012). Predicting the importance of freedom of speech and the perceived harm of hate speech. *Journal of Applied Social Psychology, 42*(6), 1353–1375. doi:10.1111/j.1559-1816.2012.00902.x.

Engelmann, I., & Marzinkowski, H. (2017, May). Most regularly used online news outlets (widest reach per country) and their platform characteristics. Presentation at the International Communication Association pre-conference "Comments, anyone? Multidisciplinary approaches for analyzing online user comments across news and other content formats", May 25, 2017, San Diego, CA, USA.

Eranti, V., & Lonkila, M. (2015). The social significance of the Facebook like button. *First Monday*, *20*(6). doi:10.5210/fm.v20i6.5505

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. doi:10.1037/a0023304

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. doi:10.1080/19331681.2014.997416

Gibbs, J. L., Kim, H., & Ki, S. (2016). Investigating the role of control and support mechanisms in members' sense of virtual community. *Communication Research*, advance online publication. doi:10.1177/0093650216644023

Goodman, E. (2013). Online comment moderation: emerging best practices. Retrieved from http://www.wan-ifra.org/reports/2013/10/04/online-comment-moderation-emerging-best-practices

Graham, T., & Witschge, T. (2003). In search of online deliberation: Towards a new method for examining the quality of online discussions. *Communications*, *28*(2), 173–204. doi:10.1515/comm.2003.012

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

Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society.* Cambridge, MA: Polity.

Hölig, S. & Hasebrink, U. (2015). Reuters digital news survey 2015. Retrieved from http://www.hans-bredow-institut.de/webfm_send/1095

Hong, J., & Park, H. S. (2011). User ratings and willingness to express opinions online. *International Journal of Marketing Studies*, *3*(2), 2–15. doi:10.5539/ijms.v3n2p2

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

Ikeda, K., & Boase, J. (2011). Multiple discussion networks and their consequence for political participation. *Communication Research*, *38*(5), 660–683. doi:10.1177/0093650210395063

Imhoff, R., & Recker, J. (2012). Differentiating Islamophobia: Introducing a new scale to measure Islamoprejudice and secular Islam critique. *Political Psychology, 33*(6), 811–824. doi:10.1111/j.1467-9221.2012.00911.

Jensen, E. (2016, August 17). NPR website to get rid of comments. *NPR*. Retrieved from http://www.npr.org/sections/ombudsman/2016/08/17/489516952/npr-website-to-get-ridof-comments

Kenski, K., Coe, K., & Rains, S. A. (2017). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research,* advance online publication. doi:10.1177/0093650217699933

Kim, N. (2016). Beyond rationality: The role of anger and information in deliberation. *Communication Research*, *43*(1), 3–24. doi:10.1177/0093650213510943

Kühl, E. (2015). Melden, anzeigen, anprangern [Flagging, reporting, denouncing]. *ZEIT Online*. Retrieved from http://www.zeit.de/digital/internet/2015-08/facebook-kommentare-hass-hetze-anzeige

Levine, M., & Crowther, S. (2008). The responsive bystander: how social group membership and group size can encourage as well as inhibit bystander intervention. *Journal of Personality and Social Psychology*, *95*(6), 1429–1439. doi:10.1037/a0012634

Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, *26*(6), 1536–1546. doi:10.1016/j.chb.2010.05.023

Mutz, D. C. (2008). Is deliberative theory a falsifiable theory? *Annual Review of Political Sciences*, *11*, 521–538.

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

Naab, T. K. (2012). The relevance of people's attitudes towards freedom of expression in a changing media environment. *ESSACHESS – Journal for Communcation Studies, 5*(1). Retrieved from http://www.essachess.com/index.php/jcs/issue/view/10

Naab, T. K. (2016). Der Sanktionsbedarf von Facebook-Inhalten aus Sicht von NutzerInnen und seine Determinanten [The need for penalties against Facebook content and its determinants. A users' perspective]. *Medien & Kommunikationswissenschaft, 64*(1), 56-73. doi:10.5771/1615-634X-2016-1-56.

Naab, T. K., Kalch, A., & Meitz, T. G. K. (2016). Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on bystander behavior. *New Media & Society*. Advance online publication. doi:10.1177/1461444816670923

Ng, E. W. J., & Detenber, B. H. (2005). The impact of synchronicity and civility in online political discussions on perceptions and intentions to participate. *Journal of Computer-Mediated Communication*, *10*(3). doi:10.1111/j.1083-6101.2005.tb00252.x

Noci, J. D., Domingo, D., Masip, P., Micó, J. L., & Ruiz, C. (2010). Comments in news, democracy booster or journalistic nightmare: Assessing the quality and dynamics of citizen debates in Catalan online newspapers. *International Symposium on Online Journalism*, *2*(1), 46–64.

Oetzel, J., Ting-Toomey, S., Chew-Sanchez, M. I., Harris, R., Wilcox, R., & Stumpf, S. (2003). Face and facework in conflicts with parents and siblings: A cross-cultural comparison of Germans, Japanese, Mexicans, and U.S. Americans. *Journal of Family Communication*, *3*(2), 67–93. doi:10.1207/S15327698JFC0302_01

Ogan, C., Willnat, L., Pennington, R., & Bashir, M. (2014). The rise of anti-Muslim prejudice: Media and Islamophobia in Europe and the United States. *International Communication Gazette, 76*(1), 27–46. doi:10.1177/1748048513504048

Oozeer, A. (2014). Internet and social networks: freedom of expression in the digital age. *Commonwealth Law Bulletin*, 40(2), 341–360. doi:10.1080/03050718.2014.909129

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

Reich, Z. (2011). User comments. The transformation of participatory space. In J. B. Singer, A. Hermida, D. Domingo, A. Heinonen, S. Paulussen, T. Quandt, . . . M. Vujnovic (Eds.), *Participatory journalism: Guarding open gates at online newspapers* (pp. 6–117). Chichester, UK: Blackwell.

Rossini, P. (2017). It's not deliberation (but that's okay)! Uncivil discourse, intolerance, and political talk. Presentation at the annual conference of the International Communication Association, May 26–29, 2017, San Diego, CA, USA.

Rowe, I. (2014). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society, 18*(2), 121–138.

Ruiz, C., Domingo, D., Mico, J. L., Diaz-Noci, J., Meso, K., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics*, *16*(4), 463–487. doi:10.1177/1940161211415849

Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, *8*(1), 18–33. doi:10.1080/17512786.2013.813194

Sarapin, S., & Morris, P. (2014). When "like"-minded people click: Facebook interaction conventions, the meaning of "speech" online, and Bland v. Roberts. *First Amendment Studies*, *48*(2), 131–157. doi:10.1080/21689725.2014.962557

von Sikorski, C. (2016). The effects of reader comments on the perception of personalized scandals: Exploring the roles of comment valence and commenters' social status. *International Journal of Communication*, *10*, 4480–4501.

Singer, J. B. (2009). Separate spaces: Discourse about the 2007 Scottish elections on a national newspaper web site. *The International Journal of Press/Politics*, *14*(4), 477–496. doi:10.1177/1940161209336659

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

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of Public Deliberation*, *3*(2), 1–35.

Stroud, N. J., Muddiman, A., & Scacco, J. M. (2016). Like, recommend, or respect? Altering political behavior in news comment sections. *New Media & Society*. Advance online publication. doi:10.1177/1461444816642420

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. doi:10.1111/jcc4.12104

Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility? *Communication Monographs*, *83*(4), 535–556. doi:10.1080/03637751.2016.1201207

Sukumaran, A., Vezich, S., McHugh, M., & Nass, C. (2011). Normative influences on thoughtful online participation. In D. Tan (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3401–3410). New York, NY: ACM.

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. doi:10.1080/15205430903225571

Zick A., Küpper B., & Hövermann A. (2011). Intolerance, prejudice and discrimination. A European report. *Friedrich Ebert Foundation.* Retrieved from http://library.fes.de/pdf-files/do/07908-20110311.pdf

Ziegele, M. (2016). *Nutzerkommentare als Anschlusskommunikation. Theorie und qualitative Analyse des Diskussionswerts von Online-Nachrichten* [User comments as follow-up communication. Theory and qualitative analysis of the discussion value of online news]. Wiesbaden, Germany: Springer.

Ziegele, M., Johnen, M., Bickler, A., Jakobs, I., Setzer, T., & Schnauber, A. (2013). Männlich, rüstig, kommentiert. Einflussfaktoren auf die Aktivität kommentierender Nutzer von Online-Nachrichtenseiten [Male, hale, comments? Factors influencing the activity of commenting users on online news websites]. *Studies in Communication | Media, 2*(1), 67–114. doi:10.5771/2192-4007-2013-1-67

Appendix

Appendix 1. Screenshot of the comment section

Folter für die Meinungsfreiheit

Saudi-Arabien steht wegen Verletzung der Menschenrechte international in der Kritik. Nach der Verurteilung des Bloggers Raif Badawi zu einer Folterstrafe von 1.000 Stockhieben versammelten sich Menschen aus aller Welt zu Protesten auf der Straße.

International wächst die politische Kritik an der strikten Auslegung des Islams im Fall des liberalen Bloggers. 130 Politiker des Europäischen Parlaments wandten sich in einem Brief persönlich an den saudischen König Salman ibn Abd al-Aziz und forderten die Freilassung Badawis. Die ersten 50 Stockhiebe wurden Raif Badawi bereits Anfang des Jahres öffentlich erteilt. Die Wiederaufnahme der Folter wurde aufgrund von Badawis schlechtem Gesundheitszustand verschoben. So lautet die offizielle Begründung, Medienberichten zufolge kann das Aussetzen der Strafe auch als Erfolg des internationalen Drucks gewertet werden.

Demonstranten fordern in London vor der Botschaft Saudi-Arabiens die Freilassung des Bloggers Raif Badawi. (Foto: dpa)

An der Haftstrafe des Bloggers hält Saudi-Arabien jedoch weiterhin fest. Das saudische Gericht hat das Strafmaß mit zehn Jahren Haft und 1.000 Stockhieben bestätigt. Zusätzlich muss Raif Badawi eine Geldstrafe von etwa 240.000 Euro entrichten, wie die saudische Zeitung Okaz berichtete. Das sei ein „schwarzer Tag für die Meinungsfreiheit", so die Menschenrechtsorganisation Amnesty International. Der 31-Jährige wurde bereits vor drei Jahren festgenommen, nachdem er auf seinem Blog „Free Saudi Liberals" für die Meinungsfreiheit eingetreten war und das Königreich Saudi-Arabien für seine radikale Auslegung des Islams scharf kritisierte. Badawi wurde mit der Begründung, den Islam beleidigt zu haben, verhaftet. Die Folterstrafe bedeutet medizinischen Einschätzungen zufolge einen Tod auf Raten.

Badawi wird Apostasie, der „Abfall vom Islam" vorgeworfen, was mit harten und strikten Strafen geahndet wird. So sieht es die Staatsdoktrin Saudi-Arabiens vor, die neben dem Wahabismus auch einer radikalen Auslegung des islamischen Gesetzes, der Scharia, folgt. Diese schreibt unter anderem eine strenge Geschlechtertrennung in der Öffentlichkeit vor, ein Autofahrverbot für Frauen und bestraft Rebellion mit dem Tod. Die islamische Religionspolizei vor Ort sorgt dafür, dass die radikalen Gesetze eingehalten werden, Zuwiderhandlungen werden mit schweren körperlichen Strafen wie Amputationen oder Auspeitschungen geahndet – ein Strafmaß, das auch vor Ausländern oder Minderjährigen keinen Halt macht. Allein in den ersten fünf Monaten im Jahr 2015 fanden laut Zählungen der Menschenrechtsorganisation Human Rights 85 öffentliche Hinrichtungen statt, beinahe so viele wie im gesamten Jahr 2014. Die Organisation "Reporter ohne Grenzen" setzte Saudi-Arabien zudem in der Rangliste der Pressefreiheit auf Platz 164 der 180 geprüften Länder.

Badawis Frau Ensaf Haider fand mit den gemeinsamen Kindern Asyl in Kanada, von wo aus sie sich für die Freilassung ihres Mannes stark macht. Die Immigrationsministerin des kanadischen Québec, Kathleen Weil, hat sich aus humanitären Gründen auch zur Aufnahme von Raif Badawi bereit erklärt. Trotz Verurteilung tritt Badawi weiter für seine liberalen Einsichten ein. Im April veröffentlichte der Blogger sein Buch „1000 Peitschenhiebe. Weil ich sage, was ich denke". Das Buch erschien bereits in deutscher, französischer und englischer Sprache. Das Vorwort dazu diktierte Badawi heimlich aus dem Gefängnis. Darin schreibt er unter anderem: „All dieses grausame Leid ist mir und meiner Familie nur deswegen widerfahren, weil ich meine Meinung ausgedrückt habe. All das war der Preis für jeden Buchstaben in diesem Buch."

Kommentarbereich

*Respektvolles Miteinander und der Einsatz für eine konstruktive und kritische Auseinandersetzung sind uns wichtig! Gerade polarisierende und gesellschaftskritische Themen laden immer wieder zu angreifenden Kommentaren ein. Deshalb lesen Sie bitte zuerst unsere Netiquette. Zeigen Sie, wenn Ihnen ein Kommentar besonders gefällt oder Sie diesen nicht gut finden. Unsere **"Positiv"**- und **"Negativ"-Buttons** geben Ihnen die Möglichkeit, eine persönliche Bewertung abzugeben. Wenn Ihnen ein Kommentar auffällt, der andere Nutzer oder Dritte diskriminiert, verletzt oder provoziert und hier nicht stehen sollte, können Sie ihn uns melden. Nutzen Sie hierfür einfach den **"Melden"-Button**. Die Redaktion wird sich darum kümmern und gegebenenfalls Maßnahmen ergreifen. Nutzen Sie das Antwortfeld, um Zustimmung oder Kritik zu äußern. Denn nur so kann ein konstruktiver Meinungsaustausch entstehen.*

Leserkommentare

1. Nutzer Pj8

Man, was haben wir für ein Glück, dass wir in Deutschland leben!

Antworten Sie direkt auf diesen Kommentar. (Ihr Beitrag wird gespeichert, sobald Sie die Seite verlassen.)

2. Nutzer Fb5

Das Buch ist sehr spannend und interessant. Man bekommt Einblicke in den kulturellen und politischen Hintergrund, die mir vollkommen neu waren.

Antworten Sie direkt auf diesen Kommentar. (Ihr Beitrag wird gespeichert, sobald Sie die Seite verlassen.)

3. Nutzer Hk6

Was für ein schreckliches, menschenverachtendes Verfahren! Diese Menschen töten, wenn jemand seine Meinung sagt. Die verantwortlichen Politiker sollten von der Welt verfolgt und ihnen der Prozess gemacht werden. Unvorstellbar, dass viele arabische Muslime da einfach mitspielen, schweigen und wegsehen. Ich würde jeden einzelnen Araber, der dazu nichts sagt, vor Gericht stellen.

Antworten Sie direkt auf diesen Kommentar. (Ihr Beitrag wird gespeichert, sobald Sie die Seite verlassen.)

Artikel kommentieren

Verfassen Sie hier einen Kommentar zum Artikel. (Ihr Beitrag wird gespeichert, sobald Sie die Seite verlassen.)

Appendix 2. Header and teaser of the stimulus article

Torture for freedom of opinion

Saudi Arabia provokes international criticism for violating human rights. When the blogger Raif Badawi was convicted and sentenced to 1000 lashes, people protested against the punishment all over the world.

Appendix 3. Stimuli comments

Uncivil and impolite comment

What a fucked up, perverse treatment! These bastards kill, when someone speaks out. The shitty dictator pigs should be chased down by the world, stoned and executed publicly. Unbelievable, how many filthy Muslims as always don't look through their yashmak. I would whip and shoot every single of these sons of bitches who doesn't say a word against that.

Uncivil and polite comment

What a terrible, inhuman treatment! These people murder, when someone speaks out. The politicians in charge should be chased down by the world, stoned and executed publicly. Unbelievable, how many Arab Muslims as always don't look through their yashmak. I would whip and shoot every single Arab who doesn't say anything against that.

Civil and impolite comment

What a fucked up, perverse treatment! These bastards kill, when someone speaks out. The shitty dictator pigs should be pursued by the world and taken to court. Unbelievable, how many filthy Muslims join, keep silent and look away. I would take every single of these sons of bitches to court who doesn't say anything against that.

Civil and polite comment

What a terrible, inhuman treatment! These people murder, when someone speaks out. The politicians in charge should be pursued by the world and taken to court. Unbelievable, how many Arab Muslims join, keep silent and look away. I would take every single Arab to court who doesn't say anything against that.