The impact of giving feedback in online discussions: effects of evaluative reply comments on the authors of evaluated user comments

Teresa K. Naab

Angaben zur Veröffentlichung / Publication details:

Naab, Teresa K. 2022. "The impact of giving feedback in online discussions: effects of evaluative reply comments on the authors of evaluated user comments." *Journal of Media Psychology* 34 (6): 334–47. https://doi.org/10.1027/1864-1105/a000330.

Original Article

The Impact of Giving Feedback in Online Discussions

Effects of Evaluative Reply Comments on the Authors of Evaluated User Comments

Teresa K. Naab

Department of Media, Knowledge and Communication, University of Augsburg, Germany

Abstract: In online discussions, users often evaluate comments from other users. On the basis of face theory, the present study analyzed the effects of evaluative replies on the evaluated comment authors. The investigation complements existing research, which has mainly focused on effects of comments on uninvolved readers. In the experimental study presented here, disapproving evaluations provoked negative and less positive emotions, and the evaluated authors were less willing to participate in the online discussion further. The authors' perception of face threat mediated these effects. The results contribute to face theory in computer-mediated interactions and to our understanding of online discussions with dissonant standpoints.

Keywords: user comments, face theory, emotion, participation, experiment

In online discussions below social media postings, users can express their views and evaluate comments made by others. Various studies reveal that feedback comments that support or criticize other comments influence uninvolved readers. For example, they affect readers' perception of the initial comment, their attitudes toward issues, and their participation in the discussion (e.g., Jin et al., 2015; Stroud et al., 2015; Ziegele & Jost, 2020).

By contrast, the effects of feedback comments on the authors of the initial comments have barely been the subject of investigation. This is surprising, since many feedback authors do not intend to only affect comment readers. They are also motivated to correct or encourage other comment authors; they aim to reinforce, signal affinity, provoke thinking, or ostracize (Springer et al., 2015; Ziegele, 2016). For this reason, the present article asks: How do evaluative reply comments in online discussions affect the perceptions, emotions, and willingness to participate of the authors of the evaluated comments?

To understand the effects of evaluative replies on the contributors in online discussions, the article builds on face theory (Goffman, 1967). In an experiment, I tested the effects of evaluative reply comments on the evaluated comment authors' emotions and willingness to participate further. Thereby, the study is among the first to differentiate the effects of not only negative evaluations (disapproving and mixed replies) but also approving evaluations. I tested the mediating effect of the authors' perceived threat to their positive and their negative face, that is, to their desire to be appreciated by others and their wish for autonomy. Finally, I tested the indirect effect of evaluative replies on the willingness to participate via emotions.

The results expand our knowledge on the applicability of face theory to computer-mediated contexts. They help to understand the process of norm negotiation in comment sections because commenters give evaluative feedback to others to indicate their approval of positions and persons and to sanction others. Examining the effects of differently valenced responses helps to understand the users' perception of online discussions and why they (do not) further participate in discussions when others offer dissonant perspectives.

Face Threats of Evaluations

In social interactions, people are concerned about their face, their positive social image created in the context of social interaction (Cupach & Metts, 1994; Goffman, 1967; Oetzel et al., 2000). They desire *positive face;* this is a wish for a self-image to be appreciated and approved by others. They also desire *negative face*; this is a wish to be autonomous and not imposed on by others. Social interactions that comply with these wishes are perceived as polite and

indicate smooth social relationships. Other interactions, among them negative evaluations, deny these desires. They hold the risk of a so-called face threat to positive or negative face (Brown & Levinson, 1987).

Although face theory has originally been tested in face-to-face interactions, it is an overarching framework to understand social interactions in various contexts, including computer-mediated communication (Brett et al., 2007). In both computer-mediated and face-to-face situations, people are concerned about their face (Bedijs et al., 2014). This is underlined by the finding that people expect politeness (i.e., support for their face; Brown & Levinson, 1987) in online interactions in a similar way to offline interactions (Reeves & Nass, 1996). People also react to face-threatening online interactions in line with the assumptions of face theory (e.g., Chen, 2015).

In online discussions, commenters often evaluate other comment authors. The face-threatening potential of such evaluations varies with their valence. Three broad types of evaluations are differentiated in the following: disapproving, mixed, and approving. In particular, *disapproving* evaluations have the potential to be hurtful (Brown & Levinson, 1987; Vangelisti, 1994; Zhang & Stafford, 2008). For example, commenters receiving a reply depreciating their viewpoint or doubting the quality of their argument may feel devalued (threat to their positive face). They may also perceive an intrusion in their autonomy of participating in the discussion and pressure to a future course of action, such as withdrawing their statement (threat to their negative face). Feedback-givers who use this type of evaluation accept a threat to the face of the feedback-receivers. In politeness research, such purely negative comments indicate a most impolite response (Brown & Levinson, 1987). Communication scholars use a differentiation that overlaps with politeness research: With the term "impolite" they mainly refer to expressions and behaviors that do not acknowledge the etiquette basics (Papacharissi, 2004, p. 280). Complementarily, the term "uncivil" refers to expressions and behaviors that violate democratic norms (Papacharissi, 2004). It often includes insults, vulgarity, or discrimination (Coe et al., 2014).[1]

In *mixed* or ambivalent evaluations, feedback-givers do not plainly oppose the feedback-receivers. They take a less decided standpoint, consider the others' interests along with their own, and try to integrate perspectives (Oetzel et al., 2001). They honor the receivers' positive face (e.g., through appreciating some of their arguments) or negative face (e.g., through accepting their freedom to choose sides or express their view) at least to some extent.

The third type of evaluations considered in this study is *approval*. These purely positive evaluations create no face threat to the receivers but fully advocate the interaction partner. Note that civil responses are not necessarily approving but could also be civilly formulated disagreement (while approval is mostly civil and polite).

Evaluations can vary in their degree of reasoning and argumentation. While some user comments present well-reasoned arguments and facts (Graham & Wright, 2015), several studies suggest that a relevant number of comments are not based on evidence (Loveland & Popescu, 2011; Ziegele et al., 2014). Therefore, this study focused on the face-threatening effects of *merely valenced reply comments*. These approve or disapprove of an idea expressed in a comment, its style, or characteristics of its author, yet they are not substantiated by arguments. In the words of Price et al. (2006), they do not have informational influence. Still, they transport normative information about the expectations of the feedback-giver and can influence the receiver (Price et al., 2006).

Affective Outcomes

Perceived face threats can lead to emotional consequences. Emotions are here understood as short-lived and context-specific states (Nabi, 2010). Emotions of positive valence reflect being enthusiastic, alert, and active. Negative emotions reflect feelings of subjective distress (Watson et al., 1988). Goffman (1967) contends that individuals are emotionally invested and attached to their faces. Face-threatening messages have been found to increase negative emotions. They hurt (Zhang & Stafford, 2008) and elicit anger, embarrassment, and anxiety (Cupach & Carson, 2002; Kennedy-Lightsey, 2010; Lin & Yamaguchi, 2011; Willer & Soliz, 2010). This holds for threats to negative as well as positive face. Rains (2013) shows in a meta-analysis that freedom threats (which are threats to negative face) can cause negative emotions such as anger. Research on online ostracism indicates that people who were excluded (i.e., those who experience a threat to their negative face) experienced negative emotions (Zadro et al., 2004). Similarly, users who were criticized or rejected by strangers in social media (i.e., positive face threat) experienced more negative feelings than users who received appreciating replies (Chen, 2015). However, Chen and Lu (2017) do not find support for their hypothesis that participants who received disagreeing replies on their comments (i.e., positive face threat) experience more negative emotions than those who did not receive any reply. It should

[1] Scholars use the term "disagreement" to denote disapproval of an argument or a contrasting standpoint. It is not necessarily uncivil (Chen & Lu, 2017), yet it might be perceived as impolite (Brown & Levinson, 1987; on disagreement, see further in this section).

be noted, however, that they did not compare the emotions of a group who received an agreeing or otherwise positive response.

By contrast, face-giving interactions, such as confirmations of affinity and trust as well as conforming arguments or approving evaluations, are associated with positive emotions (Brett et al., 2007). Jucks et al. (2016) support the positive effects of ambivalent and polite teacher feedback in an online learning environment compared to frank and disapproving evaluations. However, Chen (2015) does not find significantly more positive emotions after nonaversive replies compared to criticism and rejection in social media.

From the reported evidence on the influence of face threats on negative and positive emotions and the elaboration of the face-threatening character of different types of evaluations, it is postulated:

Hypothesis 1 (H1): Disapproving evaluative replies lead to more negative emotions of the evaluated authors than approving or mixed replies.

Hypothesis 2 (H2): Disapproving evaluative replies lead to less positive emotions of the evaluated authors than approving or mixed replies.

It is assumed that the valence of the evaluation determines positive and negative emotions because it influences the perceived threat to positive and negative face.

Hypothesis 3 (H3): The effect of the valence of the evaluation on negative emotions is mediated through (a) perceived threat to positive face and (b) perceived threat to negative face.

Hypothesis 4 (H4): The effect of valence of the evaluation on positive emotions is mediated through (a) perceived threat to positive face and (b) perceived threat to negative face.

Willingness to Participate Further

Face-threatening evaluations can have behavioral effects. People perceive face-threatening interactions as unsatisfactory (Cupach & Carson, 2002). As a consequence, they might avoid further conflict (Oetzel et al., 2001). Research on discussions in cross-cutting social networks indicates that counter-attitudinal contributions (similar to disapproving evaluations) can undercut the participants' willingness to express themselves in the conversation (Lu & Gall Myrick, 2016; McDevitt et al., 2003; Mutz, 2002; Wojcieszak & Price, 2012). This is also supported by Munger (2017), who, arguing along social norm theory, shows that disapproving responses can reduce Twitter authors' future racist tweets.

However, there are also reasons to anticipate greater participation after negative evaluations. Participants in an online deliberation experiment seemed to tolerate disagreement, "because they view it as necessary to achieve political ends" (Stromer-Galley & Muhlberger, 2009, p. 186). This is in line with face-negotiation theory (Oetzel et al., 2000), which suggests that people should not necessarily avoid conflicts. Instead, people might, "apologize, compromise, consider the other, [engage in] private discussion, and talk about the problem. This factor emphasizes a mutual concern for both self-face and other-face" (Oetzel et al., 2000, p. 413). This means evaluations that are not fully positive should lead to further participation and a willingness to resolve the issue.

People may also participate after being evaluated; they may argue, self-defend, and retaliate. This is supposed to restore one's own face by damaging the others' (Brett et al., 2007; Cupach & Metts, 1994; Oetzel et al., 2001). For example, participants who were rejected and criticized by a social network group retaliated against the group by sending aggressive emoticons (Chen, 2015). Commenters who received (uncivil or civil) disagreeing comments had more aggressive intentions compared to commenters who did not receive a comment (Chen & Lu, 2017). Those who received uncivil (i.e., more face-threatening) disagreement were even more likely to send aggressive emoticons than receivers of civil disagreement or no response (Chen & Lu, 2017; note that the authors did not compare an agreeing response). Hwang et al. (2018) propose that such a defensive reaction may be particularly likely when the opposing statement is perceived as hostile. Cheng et al. (2014) provide evidence from content analyses that commenters who receive more negative feedback show increased activity later and produce comments of lower quality. In sum, and contrary to the arguments in the previous paragraph, these considerations suggest greater participation after disapproving evaluations.

Given the adversarial arguments, the study poses two research questions:

Research Question 1 (RQ1): How does the valence of evaluative replies affect the willingness to further participate of the evaluated authors?

Research Question 2 (RQ2): Is this effect mediated through a perceived threat to positive and negative face?

There is reason to consider that the effect of evaluative replies on willingness to participate further is also mediated through emotions. Emotions can trigger interpersonal communication (Rimé, 2009). They are an important predictor

of political social media participation in general (Heiss, 2020) and people's willingness to participate in comment sections in specific (Diakopoulos & Naaman, 2011). Negative affective involvement increases users' willingness to reply to others' comments (Ziegele et al., 2018). Affective intelligence theory suggests that negative emotions, in particular anxiety, can mobilize users through elaboration. This can result, for example, in commenting behavior to prevent disliked outcomes (Heiss et al., 2019; Marcus et al., 2000). Anger, too, has been shown to encourage political discussions to combat one's beliefs (Lu & Gall Myrick, 2016).

Positive emotions can also trigger commenting behavior because they can reinforce existing behavior (Marcus et al., 2000) and make users more open to engaging with positive content (Berger & Milkman, 2012). Yet, positive emotions could also inhibit more effortful behavior because people may infer that sufficient progress has been made (Orehek et al., 2011).

Following from this, I ask:

Research Question 3 (RQ3): Is the effect of the valence of evaluative replies on the evaluated authors' willingness to further participate mediated through negative and positive emotions?

Reference of the Evaluation

Beyond the valence of an evaluation (disapproving, mixed, approving), the content of assessments can vary in many further aspects. These determine the severity of the face-threatening act (Brown & Levinson, 1987). The present study focused on merely evaluative replies and kept the potential influence of arguments constant. A further determinant that is assumed to impact the severity of the face-threatening character of evaluations is the reference category of the evaluations: (1) Evaluations can be directed at the behavior or position rather than the person (Smith et al., 1998). Negative evaluations of one's position often come in the form of disagreement, which is an expression of a countering or challenging viewpoint (Klofstad et al., 2013). From the perspective of deliberative democracy, references to others' positions are valuable to the exchange of ideas and – if expressed civilly (see previous section) – can contribute to deliberative opinions (Price et al., 2002). However, negative evaluations of a position can also create cognitive dissonance and challenge the evaluated person (Festinger, 1957). Such disagreement may be perceived as impolite (Brown & Levinson, 1987). (2) Evaluations can be directed at the person rather than at the behavior or position under consideration (Alberts & Driscoll, 1992). In extreme forms, these are uncivil "ad hominem attacks" (Habernal et al., 2018). Such forms of disapproval do not contribute to deliberation.

Generally, disapproving evaluations directed at a person are perceived as worse by their recipients than disapproval directed at a position (i.e., disagreement). Cupach and Carson (2002) found complaints directed at the person to be more face-threatening because they reference general and irreparable personality flaws. Stryker et al. (2016) demonstrate that people perceive attacks against the reputation or character of a person as more uncivil than attacks against arguments or highlighting mistakes. In line with this, Chen and Lu (2017) argue that uncivil disagreement poses a greater face threat than civil disagreement. It undermines the deliberative principle of equal and argument-based discussions and should pose an even greater face threat. However, they also suggest that generally, "people may be annoyed when others disagree with them, even if the discourse is civil" (p. 121), that is, when others disapprove of their position.

The reference of evaluations might interact with the effects assumed earlier. If a disapproving evaluation not only disagrees with a contributor's comment but depreciates their general character or abilities, this is a greater face threat than a disapproving evaluation of the comment content and could cause more negative emotions. By contrast, approving and mixed evaluations that address the comment author personally should be perceived as more positive. Consequently, I postulate an interaction effect on emotions:

Hypothesis 5 (H5): The reference in an evaluative reply comment moderates the effect of the valence of the evaluation on the emotions of the evaluated authors. Disapproving evaluative replies directed at the author's person lead to more negative emotions and less positive emotions than disapproving evaluative replies directed at the comment content.

Research Question 4 (RQ4): Does the reference in an evaluative reply comment moderate the effect of the valence of the evaluation on willingness to further participation of the evaluated authors?

Method

Design and Procedure

A 3 × 2 × 2 between-subject experimental design varying the valence of the evaluation (disapproving vs. mixed vs. approving), the reference of the evaluation (directed at the comment content vs. directed at the author's person), and the topic of a Facebook post (meat consumption vs. sexist advertising) was carried out. The design also included a control group that did not receive a reply comment.

This group will not be considered in the further analyses. I informed the respondents that they would take part in a study on a controversial topic. During the study, they would be able to leave comments. They were further told that other participants could read and reply to their comments. The participants were randomly shown either a Facebook post on meat consumption or advertising for a watch with a sexist slogan. Three comments with heterogeneous standpoints were posted below the post. As on Facebook, the participants could comment below the post. After reading and potentially commenting, participants were asked questions on their general Facebook participation. Then, respondents were informed that they could see the post and the discussion thread again and that other participants had probably written new comments or replied to their comment. The stimulus Facebook page then showed the initial comments and the comment of the participant. It further contained a randomized reply comment allegedly written by another participant and two additional comments not replying to the participant's comment. The additional comments were identical in all conditions. In the following, the participants were asked about their emotions, perception of the reply comment, and willingness to continue participating in the discussion. After completing the questionnaire, respondents were fully debriefed.

Participants

The experiment was conducted in Germany. Facebook users were recruited from a wide variety of Facebook pages. Participants could win gift cards. A total of 720 participants completed the full survey (65.7% female; age: M = 31.74 years, SD = 11.68; 19.8% did not hold a college entrance certificate).

The survey captured whether the participants commented on the Facebook post. Overall, 433 participants (60.1%) wrote a comment and were assigned to one of the experimental conditions or the control group without a reply. Those who wrote a comment did not differ from those who did not regarding gender (Cramer's V = .015, p = .694), education (Cramer's V = .035, p = .355), and general Facebook use frequency (1 = *never* to 7 = *several times a day*), M_{authors} = 6.06, SD_{authors} = 1.81; $M_{\text{non-authors}}$ = 5.95, $SD_{\text{non-authors}}$ = 1.87, t (718) = −0.778, p = .437. However, the authors were slightly older than the non-authors, M_{authors} = 32.59, SD_{authors} = 11.90; $M_{\text{non-authors}}$ = 30.45, $SD_{\text{non-authors}}$ = 11,24, t (636.605) = −2.449, p = .015.

The analyses only refer to participants in the experimental conditions who received a reply comment (63.8% female; age: M = 32.92 years, SD = 12.12; 18.3% did not hold a college entrance certificate). Participants were equally distributed across the experimental groups regarding age, F(5,361) = 0.427, p = .829) gender (Cramer's V = .154, p = .124), and education (Cramer's V = .146, p = .294).

Stimuli

The reply comments only varied in their valence. The disapproving reply devalued the comment/author. The mixed reply showed interest in the comment/author while at the same time indicating indecision about its quality. The approving reply advocated the comment/author. All replies touch upon positive face (appreciation/depreciation of the comment) as well as negative face ([non-] acceptance that the comment was posted in the thread). The reply comments also varied in the reference of their evaluation. In the first condition, the reply referred only to the content of the comment. In the second condition, it referred only to the abilities or lack thereof of the comment author (see Appendix in the Electronic Supplementary Material 1, ESM 1, for the replies).

To strengthen the generalizability of the results, the study varied the controversial topic of the Facebook post. The first post reported on meat consumption and advocated vegetarian nutrition for the general population. The second post advertised a watch. It used a humorous slogan that could also be perceived as sexist and evoke disapproval by stereotyping women and attacking gender equality. The comment sections below the posts contained three comments with heterogeneous perspectives on the topics. In total, 371 of the 720 participants saw the post on meat consumption; 237 of them (63.9%) posted a comment. 349 of the 720 participants saw the ad; 196 of them (56.2%) posted a comment.

Stimulus Pretest

I developed reply comments for each condition based on authentic Facebook comments. They must be typical for Facebook and serve as responses to the diverse comments expected from the participants in the main study. In a pretest, I tested whether the stimulus reply comments clearly reflected the conditions. A total of 48 respondents not involved in the main study assessed various reply comments and rated their perception. They assessed a disapproving, a mixed, and an approving stimulus reply comment. The three presented stimuli randomly referred to the comment content or the person of the author. Participants rated whether they perceived the reply comments as (1) opposing or agreeing with the comment they responded to, (2) showing interest or not showing interest in the comment they responded to, and (3) positive or negative. Additionally, the participants assessed (4) whether the replies addressed the comment content or the person of the author

and (5) whether they referred to the content or the author. They rated the five measures on a semantic differential scale from 1 (*strongly agree with item on the left*) to 6 (*strongly agree with opposing item on the right*).

The participants perceived the disapproving stimulus as more negative and opposing than the mixed and the approving stimulus. The mixed reply, in turn, was perceived as more negative and opposing than the approving reply. As intended, the disapproving reply was also perceived as showing less interest in the comment it responded to than the mixed and the approving replies. The participants evaluated the replies that were directed at the comment content as addressing their comment more than themselves and referring more to their comment than to them as a person compared to the replies that were directed at the person (see Tables EI–EIV in ESM 1 for the descriptives and statistical test results).

Measures

Data were saved on whether the participants did or did not *write a comment* beneath the post to which they were assigned. Overall, 433 participants posted a comment (60.1%); 367 were assigned to one of the experimental conditions, not the control group.

Perceived face threat of the reply comment was measured following Cupach and Carson (2002). Eight items referred to threats to positive face (e.g., "The reply comment is rude," "shows disrespect towards me"; 1 = *strongly disagree* to 7 = *strongly agree*, α = .948, M = 3.98, SD = 2.03). Four items measured perceived threat to negative face (e.g., "constrains my choices," "takes away some of my independence," α = .792, M = 2.34, SD = 1.49).

Negative and positive emotions were measured with the German version of the Positive and Negative Affect Schedule (Breyer & Bluemke, 2016). Participants rated how they felt "right now" with 10 items on negative emotions (1 = *not at all* to 5 = *extremely*, α = .899, M = 1.44, SD = 0.63) and 10 items on positive emotions (α = .877, M = 2.67, SD = 0.77).

Six items adopted from Ng and Detenber (2005) served to measure *willingness to continue participation* (e.g., "I would like to further contribute to the discussion"; 1 = *strongly disagree* to 7 = *strongly agree*, α = .827, M = 3.61, SD = 1.45).

To check for the validity of the mock setting, participants were asked to state their perception of the Facebook page. They perceived the posts as comprehensible (1 = *strongly disagree* to 7 = *strongly agree*, N = 720, M = 5.58, SD = 1.57), typical (M = 5.38, SD = 1.53), and usual for Facebook (M = 5.45, SD = 1.47). They assessed that such a comment thread could similarly be found on Facebook (M = 5.71, SD = 1.37), is typical (M = 5.54, SD = 1.47), and the user comments are realistic (M = 5.19, SD = 1.56).

Results

The data were analyzed with general linear models. In all models, valence of the evaluation, reference of the evaluation, and topic of the post served as independent factors. Besides the main effect of the three factors, the interaction effect between valence and reference was included. The dependent variables positive emotions, negative emotions, and willingness to participate further showed only low correlations (r < .3), which is why separate univariate models were run.

The valence of the evaluation influenced negative emotions (Table 1). The disapproving reply provoked significantly greater negative emotions than the mixed (diff = −0.187, p = .023, 95% CI [−0.348, −0.026]) and the approving reply (diff = −0.282, p < .001, 95% CI [−0.433, −0.131]). H1 is supported. For detailed descriptive statistics by conditions, see Table EVII in ESM 1.

The valence of the evaluation also influenced positive emotions (Table 1). The disapproving reply caused less positive emotions than the approving reply (diff = 0.260, p = .007, 95% CI [0.073, 0.448]). However, people did not experience less positive emotions after a disapproving reply compared to a mixed reply (diff = 0.139, p = .174, 95% CI [−0.061, 0.338]). H2 is partly supported. For detailed descriptive statistics by conditions, see Table EVIII in ESM 1.

Neither the reference of the evaluation nor the interaction effect between valence and reference determined negative emotions or positive emotions (Table 1). H5 is not supported.

The valence of the evaluation had a significant influence on the willingness to participate further (Table 1). The disapproving reply resulted in significantly lower willingness to continue participating than the approving reply (diff = 0.605, p = .001, 95% CI [0.253, 0.958]). Willingness to participate did not differ significantly between the disapproving and the mixed reply (diff = 0.341, p = .076, 95% CI [−0.036, 0.717]). For detailed descriptive statistics by conditions, see Table EIX in ESM 1. This answers RQ1.

Neither the reference of the evaluation nor the interaction term determined the willingness to participate further (Table 1). This answers RQ4.

The topic of the post did not significantly affect any of the dependent variables (Table 1).

It was assumed that perceived threat to positive and negative face would mediate the relationship between the evaluative reply and negative emotions, positive emotions, and willingness to participate further (H3, H4, RQ2). These hypotheses and research questions were tested with mediation analyses for multicategorical independent variables using the PROCESS macro for SPSS (Hayes & Preacher, 2014). Bootstrap standard errors and bias-corrected 95%

Table 1. General linear models explaining negative emotions, positive emotions, and willingness to participate further

	Negative emotions[a]			Positive emotions[b]			Willingness to participate further[c]		
Independent variables	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Valence of evaluation	$F(2, 360) = 7.549$	.001	.040	$F(2, 360) = 4.256$	.015	.023	$F(2, 360) = 5.831$	.003	.031
Reference of evaluation	$F(1, 360) = 1.366$	.243	.004	$F(1, 360) = 0.704$	.704	.000	$F(1, 360) = 0.649$	.421	.002
Topic	$F(1, 360) = 0.172$	.679	.001	$F(1, 360) = 0.062$	.803	.000	$F(1, 360) = 2.019$	.156	.006
Valence × Reference	$F(2, 360) = 0.844$	.431	.005	$F(2, 360) = 1.744$	.176	.010	$F(2, 360) = 0.740$	.478	.004

Note. [a]Corrected model: $F(6, 360) = 3.257$, $p = .004$, $\eta^2 = 0.051$. [b]Corrected model: $F(6, 360) = 2.021$, $p = .062$, $\eta^2 = 0.033$. [c]Corrected model: $F(6, 360) = 2.539$, $p = .020$, $\eta^2 = 0.041$.

Table 2. General linear models explaining positive and negative face threat

	Positive face threat[a]			Negative face threat[b]		
Independent variables	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Valence of evaluation	$F(2, 360) = 389.112$	.001	.684	$F(2, 360) = 61.961$	.001	.256
Reference of evaluation	$F(1, 360) = 29.338$	.001	.075	$F(1, 360) = 3.142$	.077	.009
Topic	$F(1, 360) = 0.200$	.607	.001	$F(1, 360) = 0.037$	.847	.000
Valence × Reference	$F(2, 360) = 1.851$	.159	.010	$F(2, 360) = 0.207$	.813	.001

Note. [a]Corrected model: $F(6, 360) = 137.123$, $p < .001$, $\eta^2 = 0.696$. [b]Corrected model: $F(6, 360) = 21.629$, $p < .001$, $\eta^2 = 0.265$.

confidence intervals were generated based on 10,000 bootstrap samples. Indicator coding was used to draw comparisons of the mixed and approving replies with the disapproving reply as the reference category. Perceived threat to positive and negative face served as parallel mediators. Analyses were run separately for negative emotions, positive emotions, and willingness to participate. Since the reference of the evaluation and the topic of the post did not influence the outcome variables in the previous analyses, they were not included further.[2]

In line with the assumptions, the disapproving reply caused a significantly higher threat to positive and negative face than the mixed and approving replies (Figures 1–3).

Perceived threat to positive face did not increase negative emotions. H3a is not supported. However, perceived threat to negative face increased negative emotions (Figure 1). There was also a significant indirect effect of the valence of the evaluation through negative face threat on negative emotions; the disapproving reply created higher negative face threat and therefore led to more negative emotions (Table 3). H3b is supported.

Perceived positive face threat led to less positive emotions (Figure 2). In line with the assumptions, valence of the evaluation also had a significant indirect effect through positive face threat on positive emotions; the disapproving reply created higher positive face threat and therefore led to less positive emotions (Table 3). H4a is supported. Contrary to the assumptions, perceived negative face threat led to more positive emotions. There was also a significant indirect effect of the valence of the evaluation through negative face threat on positive emotions. It indicates that the disapproving reply, compared to mixed and approving replies, leads to higher negative face threat, which leads to more positive emotions. H4b is not supported.

Perceived threat to positive face significantly decreased willingness to participate further; however, negative face threat did not (Figure 3). The valence of the evaluation also indirectly, through positive face threat, influenced willingness to participate further; the disapproving reply created higher positive face threat and therefore lowered the willingness to participate further (Table 3). This answers RQ2.

To answer RQ3, whether the influence of valence of the evaluation on willingness to participate is mediated by emotions, I conducted a further mediation analysis. Willingness to participate served as the dependent variable and negative emotions and positive emotions served as parallel mediators. Indicator coding was used to draw comparisons of the mixed and approving replies with the disapproving reply as the reference category.

The disapproving reply caused significantly higher negative emotions than the mixed and approving replies. It also triggered less positive emotions than the approving reply (Figure 4). This replicates the results of the previous analyses (Table 2). Positive emotions led to a greater willingness to further participate in the discussion. Through positive emotions, the valence of the evaluation also indirectly affected participation willingness (Table 4). However,

[2] The reference of the evaluation also did not influence the perceived positive and negative face threat, nor did it moderate the effect of the valence of the evaluation (Table 2 and Tables EV and EVI in ESM 1).

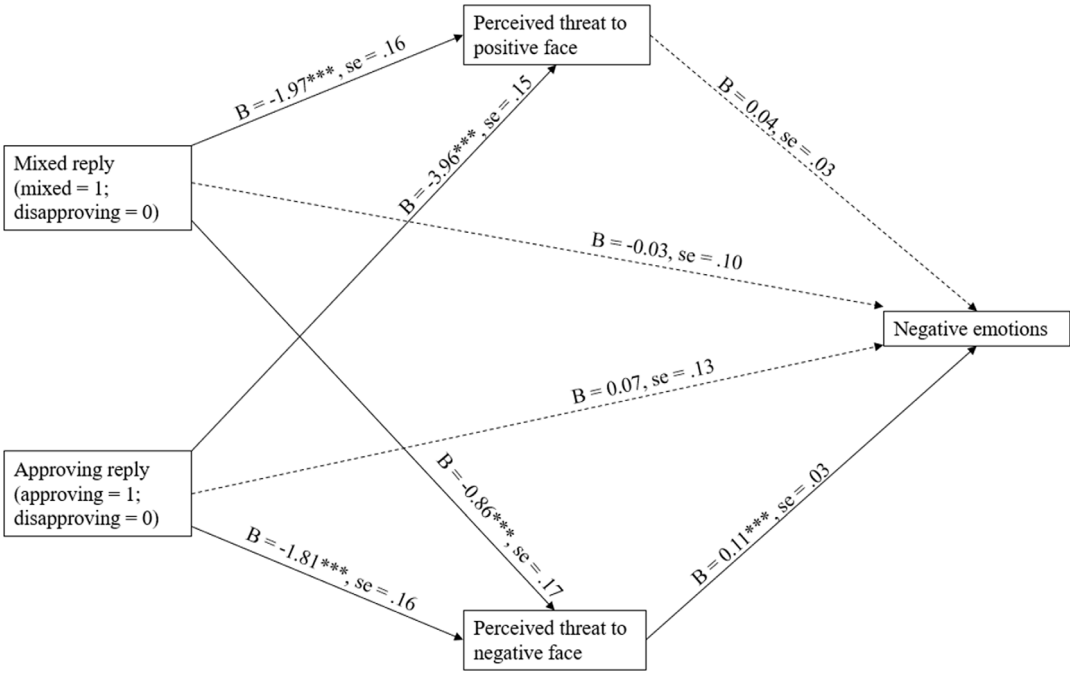

Figure 1. Direct effects of the valence of the evaluation, perceive threat to positive face, and perceived threat to negative face on negative emotions. Model summary for regression of valence of the evaluation on negative emotions: R^2 = .11, $F(4, 362)$ = 11.572, p < .001. Model summary for regression of valence of the evaluation on perceived threat to positive face: R^2 = .67, $F(2, 364)$ = 368.017, p < .001. Model summary for regression of valence of the evaluation on perceived threat to negative face: R^2 = .26, $F(2, 364)$ = 63.307, p < .001. *p < .05, **p < .01, ***p < .001. Nonsignificant paths (p < .05) are presented with dashed lines.

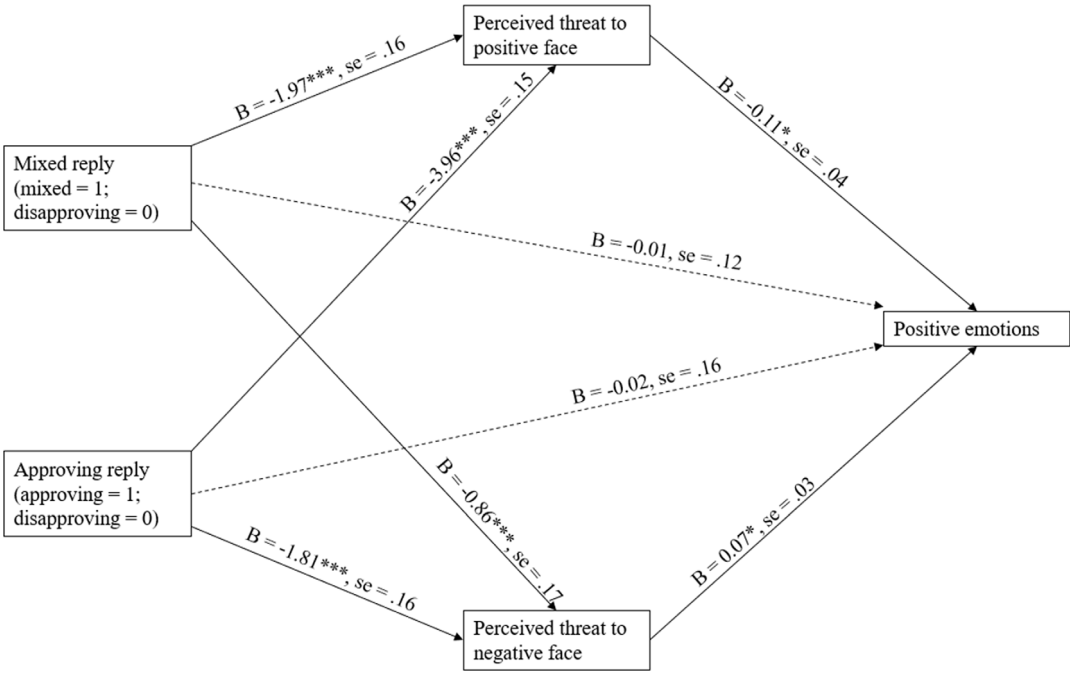

Figure 2. Direct effects of the valence of the evaluation, perceive threat to positive face, and perceived threat to negative face on positive emotions. Model summary for regression of valence of the evaluation on positive emotions: R^2 = .05, $F(4, 362)$ = 4.461, p = .002. Model summary for regression of valence of the evaluation on perceived threat to positive face: R^2 = .67, $F(2, 364)$ = 368.017, p < .001. Model summary for regression of valence of the evaluation on perceived threat to negative face: R^2 = .26, $F(2, 364)$ = 63.307, p < .001. *p < .05, **p < .01, ***p < .001. Nonsignificant paths (p < .05) are presented with dashed lines.

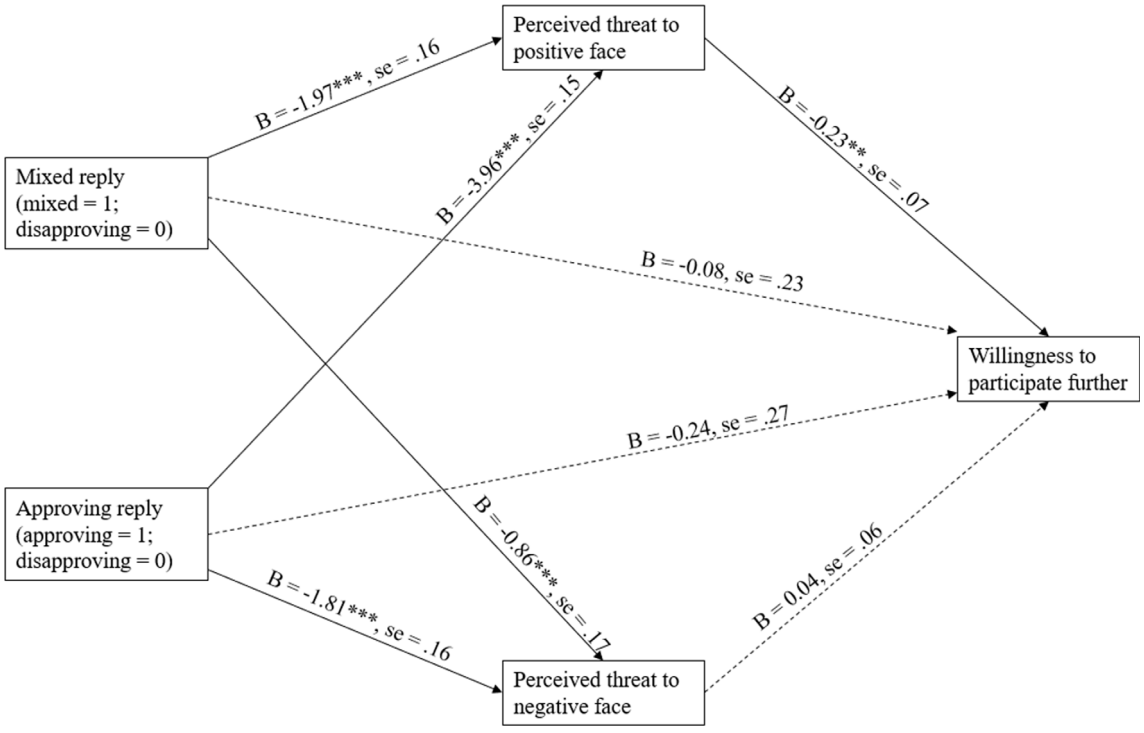

Figure 3. Direct effects of the valence of the evaluation, perceive threat to positive face, and perceived threat to negative face on willingness to participate further. Model summary for regression of valence of the evaluation on willingness to participate further: R^2 = .06, $F(4, 362)$ = 5.744, p < .001. Model summary for regression of valence of the evaluation on perceived threat to positive face: R^2 = .67, $F(2, 364)$ = 368.017, p < .001. Model summary for regression of valence of the evaluation on perceived threat to negative face: R^2 = .26, $F(2, 364)$ = 63.307, p < .001. *p < .05, **p < .01, ***p < .001. Nonsignificant paths (p < .05) are presented with dashed lines.

Table 3. Indirect effects of valence of evaluation on negative emotions, positive emotions, and willingness to participate further via perceived threat to positive face and perceived threat to negative face

		Valence of the evaluation			
		Mixed		Approving	
Dependent variable	Mediator	Indirect effect (boot *SE*)	Boot 95% CI	Indirect effect (boot *SE*)	Boot 95% CI
Negative emotions	Positive face threat	−0.10 (0.06)	[−0.23, 0.02]	−0.19 (0.12)	[−0.43, 0.04]
	Negative face threat	**−0.09 (0.04)**	**[−0.16, −0.03]**	**−0.19 (0.06)**	**[−0.31, −0.06]**
Positive emotions	Positive face threat	**0.22 (0.09)**	**[0.05, 0.39]**	**0.44 (0.17)**	**[0.10, 0.77]**
	Negative face threat	**−0.06 (0.03)**	**[−0.13, −0.01]**	**−0.13 (0.06)**	**[−0.25, −0.02]**
Willingness to participate	Positive face threat	**0.47 (0.13)**	**[0.23, 0.74]**	**0.95 (0.25)**	**[0.48, 1.45]**
	Negative face threat	−0.07 (0.06)	[−0.19, 0.02]	−0.16 (0.11)	[−0.38, 0.05]

Note. Disapproving reply serves as reference category (disapproving coded as 0; mixed/approving coded as 1). Values in bold face are significant at p < .05.

negative emotions had no effect on willingness to participate and thus did not mediate the effect of valence of the evaluation.

Discussion

Applying face theory, this study conceptualized the face-threatening character of evaluative reply comments in online discussion and its impact on the authors who receive such feedback on their comments. It aimed at expanding existing research, which has mostly examined the effects of user comments on uninvolved readers.

The findings support that valenced reply comments (which are not substantiated by reasons) indeed affect the commenters who receive feedback. Comment authors recognize disapproving replies as more threatening to their positive face, that is, as less appreciating of their comment and person compared to mixed and approving replies. They also view disapproving replies as more threatening to their negative face, which means taking away their independence compared to mixed and approving replies. The valence of

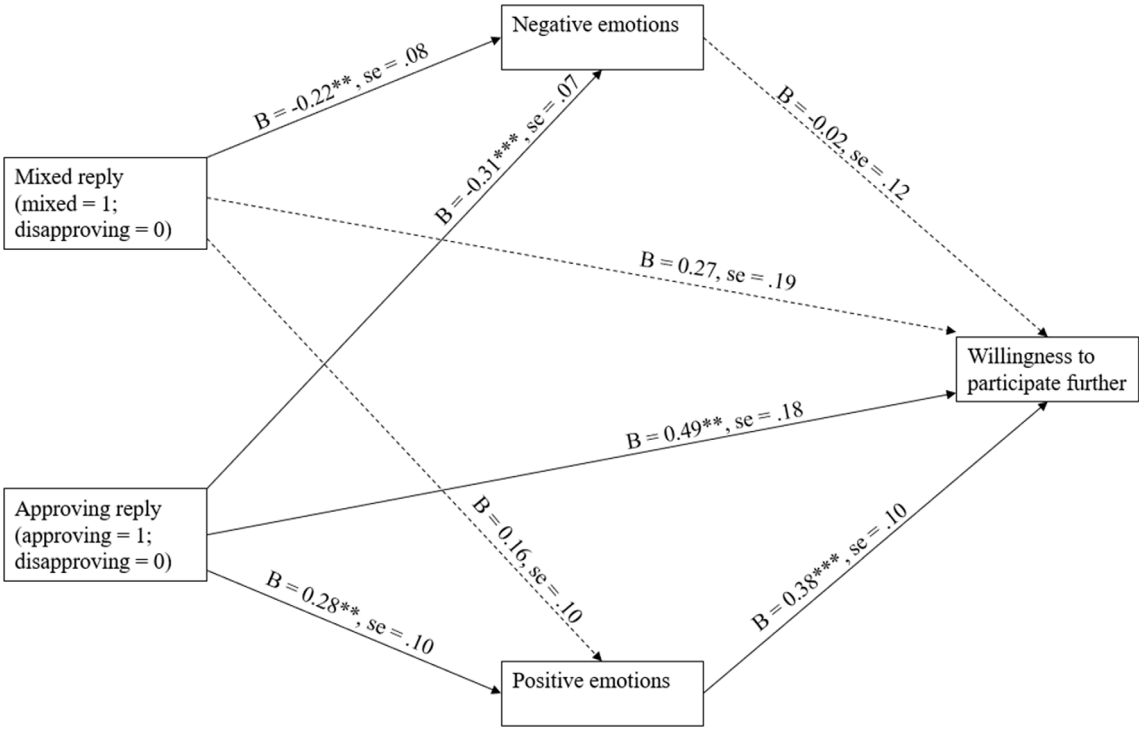

Figure 4. Direct effects of the valence of the evaluation, negative emotions, and positive emotions on willingness to participate further. Model summary for regression of valence of the evaluation on willingness to participate further: R^2 = .07, $F(4, 362)$ = 6.891, p < .001. Model summary for regression of valence of the evaluation on negative emotions: R^2 = .04, $F(2, 364)$ = 8.053, p < .001. Model summary for regression of valence of the evaluation on positive emotions: R^2 = .02, $F(2, 364)$ = 4.143, p < .001. *p < .05, **p < .01, ***p < .001. Nonsignificant paths (p < .05) are presented with dashed lines.

Table 4. Indirect effects of valence of evaluation on willingness to participate further via negative emotions and positive emotions

		Valence of the evaluation			
		Mixed		Approving	
Dependent variable	Mediator	Indirect effect (boot *SE*)	Boot 95% CI	Indirect effect (boot *SE*)	Boot 95% CI
Willingness to participate	Negative emotions	0.00 (0.03)	[−0.05, 0.06]	0.00 (0.03)	[−0.06, 0.08]
	Positive emotions	0.06 (0.04)	[−0.01, 0.16]	**0.11 (0.04)**	**[0.03, 0.20]**

Note. Disapproving reply serves as reference category (disapproving coded as 0; mixed/approving coded as 1). Values in bold face are significant at p < .05.

the evaluation explains a proper amount of variance in the perception of face threat (positive face threat: R^2 = .67, negative face threat: R^2 = .26). This supports research on face-to-face interactions, which indicates that evaluations are face-threatening acts (Brown & Levinson, 1987; Vangelisti, 1994; Zhang & Stafford, 2008).

It is an important finding that the concept of face applies to the computer-mediated context of comment sections. Compared with face-to-face interactions, participants in online discussions often interact with people they know little about. They have little status information, and power hierarchies can be lower (Dahlberg, 2001). Additionally, future interactions and the need for justifications are less likely for online discussants. However, online commenting is public and potentially reaches a broad audience; this seems to increase the desire for an advantageous self-image (Lim et al., 2012). While several communication studies have relied on face theory to understand the effects of computer-mediated communication on users (e.g., Chen & Lu, 2017), the present study is the first that tests the mediating role of both positive and negative face threat. Future studies should systematically investigate how various online settings affect the perception of face threats of evaluations, for example, by varying public availability and the relation between the conversation partners (Neubaum & Krämer, 2018).

Evaluative replies affect the emotions of the evaluated authors. Thus, a fundamental assumption in face theory applies to computer-mediated contexts. Although disapproving replies triggered both positive and negative face threats, the effect of evaluative replies on negative emotions was mediated only through negative face threats.

This means it is not the contempt and disrespect of bad evaluations but their perceived imposition that provokes negative emotions. Since the tone in many comment sections is harsh, authors might not generally expect much appreciation. Consequently, a lack of positive face might not lower their emotional state. By contrast, a perceived invasion of their right to freely express their standpoint might be more provocative of negative feelings. This points to the phenomenon of reactance, which includes anger (a negative emotion) and is triggered by freedom threats (Rains, 2013). This is in line with the present finding that threats to negative face provoke a negative emotional state.

We gained a somewhat different picture of the influence of evaluative replies on positive emotions. Only a perceived threat to positive face decreased positive feelings. A surprising finding is that a *higher* threat to negative face triggered by disapproving replies *increased* positive emotions. It seems that perceived attacks against the freedom to comment as desired made the authors more alert and active. Future studies should examine the relationship between face threats and emotions in detail, for example, through manipulating various threats to negative face.

The results also call for testing the effects of evaluations on discrete emotions. The literature indicates that social media content can influence specific positive or negative emotions. Mostly, scholars have regarded discrete emotions such as anger, aversion, and anxiety (e.g., Gervais, 2015, 2017; Lu & Gall Myrick, 2016). In line with appraisal theories, studies also find that discrete emotions can exert different effects on social media behavior (e.g., Lu & Gall Myrick, 2016; Valentino et al., 2011). Therefore, empirical tests of the influence of different types of evaluations on discrete emotions, participation behavior, and the mediating role of perceived face threat are needed. Evaluating the relationship between negative face threat and anger seems particularly fruitful; imposition and limitation of one's freedom lead to the perception of negative face threats. At the same time, when one's goals are blocked, this can trigger anger (Carver & Harmon-Jones, 2009; on reactance, Rains, 2013). The unexpected finding of the present study, that higher negative face threat increased positive emotions, calls for a distinct analysis of the effects on enthusiasm.

Willingness to participate decreases with disapproving replies. Authors experiencing such sharp evaluation tend to withdraw from the online discussion compared to approving replies. This is in line with findings of deliberation research that disagreement undercuts willingness to prolong an interaction (McDevitt et al., 2003; Mutz, 2002; Wojcieszak & Price, 2012). Here, I found no direct effects, but the influence of the valence of the evaluation was mediated through perceived threat to positive face. The more depreciating that authors perceived the reply, the less they intended to continue participation. This also mirrors the avoidance strategy reported by face-negotiation theory (Oetzel et al., 2001), which considers that people end interactions to lower the risk of further face threats. The perception of negative face threat did not lower willingness to participate. This is interesting, on the one hand, because feedback-givers who attack the negative face might intend to exclude other commenters. However, such attempts do not seem fruitful. On the other hand, it is surprising because the assumption about reactance (Rains, 2013) would suggest that feedback-receivers would counterargue as a response to a threat to their negative face. We could speculate that attacks toward the negative face might strengthen the willingness to counter the threat for some participants, while for others, it might trigger the wish to avoid further face risks. Future studies should investigate the moderating influence of individual characteristics in more detail.

Positive emotions increased users' willingness to continue participation in the discussion after receiving evaluative replies to their comments. This supports previous studies indicating that positive social media content increases engagement with the content (Berger & Milkman, 2012) and reinforces the behavior of enthusiastic users (Marcus et al., 2000). Interestingly, in contrast to several previous studies, negative emotions did not increase participation willingness. Several explanations could guide future research. The present study did not differentiate between distinct negative emotions. However, while anger could lead to combating one's beliefs in the face of disagreement, anxiety could lead to enhanced elaboration and reasoning, which does not necessarily result in further comment posting (Lu & Gall Myrick, 2016). Additionally, the reply comments did not provide arguments for their negative or mixed evaluation. This might make it difficult for the evaluated authors to respond, and negative emotions might trigger processing about one's ability and opinion instead of countering.

Unexpectedly, the reference of the evaluative reply did not influence any of the dependent variables. This might indicate a deficit of the stimulus material. The pretest suggested proper manipulation of the evaluations that were directed at the comment content and the author's person. However, in the laboratory setting, participants know that feedback givers cannot access any information about them but their comments. Thus, the participants might have related even those evaluations only to their comments, which addressed them personally in their wording. There is strong evidence that more general criticism and ad hominem attacks are more detrimental. Thus, future studies need to investigate the effect of the reference of evaluations in more natural settings that allow interaction partners to differentiate more clearly between authors and their posts.

The study also adds to existing research on the effects of user replies because it is among the first that not only compared nonpositive evaluations (disapproving, mixed) but also approving evaluations. While disapproving and mixed replies did not cause different levels of positive emotions and of willingness to participate, approving replies actually led to more positive emotions and greater willingness to participate than the two nonpositive conditions. This suggests that user feedback that is not fully positive might have equal consequences to negative feedback eventually. However, this does not hold for the effects on perceived face threat and negative emotions. Here, the study pointed to differences in disapproving and mixed evaluations.

Limitations

The results should be interpreted only in light of several limitations. First, the study used a mock Facebook page and investigated self-reported reactions to reply comments in a hypothetical situation. This procedure aimed at increasing the internal validity of the results. However, it limits ecological validity because the participants did not engage in a personalized social network site of their choice. They were also limited to one of two topics chosen by the researcher.

Additionally, the study considered only valenced replies, and the manipulated replies comprised merely three (quite extreme) types of evaluations. Future studies need to consider the effects of evaluations that come with justifications and differentiate more nuanced assessments. For example, the present study is not able to differentiate between civil and uncivil disapproving evaluations.

Despite the broad recruitment, the sample was not representative of German Facebook users. Primarily, the respondents had a higher level of education than the actual Facebook user community. Although interest in online discussions is greater among well-educated users (Hölig & Hasebrink, 2015), it would be rash to generalize the findings. Future research should examine the moderating influence of education on the perceived face threat of different types of evaluations in online communication. In face-to-face encounters, the education level of the interactants may affect the implicit hierarchies between discussion partners. By contrast, anonymous online settings provide the chance of more equal participation (Dahlberg, 2001). As power distance is a determinant of face threats (Brown & Levinson, 1987), the effects of the education level are worthy of being tested and can advance face theory. Educational level does not seem to influence the perception of uncivil online content (Kenski et al., 2020). However, it can influence conflict behavior (Bobo & Licari, 1989). Thus, an empirical test of the moderating role on future participation behavior is needed.

Conclusion

The results have implications for public discourse in comment sections. Authors do not take only evaluative replies easily but show negative emotional reactions upon disapproving responses and can lose interest in the discussion. Feedback-givers might intend such effects if they perceive a comment as in need of sanctions. Attentive moderation or engaged counter-speech by other users is required (Ziegele et al., 2020) to prevent individual or organized users from trying to suppress certain voices through attacks against the author's face. Feedback-givers are advised to consider a more integrating evaluation that considers the others' face wants. This might mean that they need to step back from extreme positions themselves. However, this could prevent reactance and contribute to a common search for consensus.

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1864-1105/a000330

ESM 1. The Electronic Supplementary Materials include the stimulus comments, the descriptives and statistical tests of the pretest, and the descriptives of the outcome variables by experimental condition.

References

Alberts, J. K., & Driscoll, G. (1992). Containment versus escalation: The trajectory of couples' conversational complaints. *Western Journal of Communication, 56*, 394–412. https://doi.org/10.1080/10570319209374425

Bedijs, K., Held, G., & Maaß, C. (Eds.). (2014). *Face work and social media*. Lit.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Bobo, L., & Licari, F. C. (1989). Education and political tolerance: Testing the effects of cognitive sophistication and target group affect. *Public Opinion Quarterly, 53*(3), 285–305. https://doi.org/10.1086/269154

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

Breyer, B., & Bluemke, M. (2016). *Deutsche Version des Positive and Negative Affect Schedule PANAS* [German version of the Positive and Negative Affect Schedule PANAS]. https://www.gesis.org/zis

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

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin, 135*(2), 183–204. https://doi.org/10.1037/a0013965

Chen, G. M. (2015). Losing face on social media: Threats to positive face lead to an indirect effect on retaliatory aggression through negative affect. *Communication Research, 42*(6), 819–838. https://doi.org/10.1177/0093650213510937

Chen, G. M., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media, 61*(1), 108–125. https://doi.org/10.1080/08838151.2016.1273922

Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2014). *How community feedback shapes users behavior*. https://cs.stanford.edu/people/jure/pubs/disqus-icwsm14.pdf

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

Cupach, W. R., & Carson, C. L. (2002). Characteristics and consequences of interpersonal complaints associated with perceived face threat. *Journal of Social and Personal Relationships, 19*(4), 443–462. https://doi.org/10.1177/0265407502019004047

Cupach, W. R., & Metts, S. (1994). *Facework*. Sage.

Dahlberg, L. (2001). Computer-mediated communication and the public sphere: A critical analysis. *Journal of Computer-Mediated Communication, 7*(1). https://doi.org/10.1111/j.1083-6101.2001.tb00137.x

Diakopoulos, N., & Naaman, M. (2011). Towards quality discourse in online news comments. In P. J. Hinds, J. C. Tang, J. Wang, J. Bardam, & N. Ducheneaut (Eds.), *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 133–142). ACM.

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

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

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

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

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

Habernal, I., Wachsmuth, H., Gurevych, I., & Stein, B. (2018). *Before name-calling: Dynamics and triggers of ad ominem fallacies in web argumentation*. Proceedings of NAACL-HLT 2018, New Orleans, LA.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *The British Journal of Mathematical and Statistical Psychology, 67*(3), 451–470. https://doi.org/10.1111/bmsp.12028

Heiss, R. (2020). The affective foundation of political opinion expression on social media: A panel analysis. *International Journal of Public Opinion Research, 15*(1), 1. https://doi.org/10.1093/ijpor/edaa009

Heiss, R., Schmuck, D., & Matthes, J. (2019). What drives interaction in political actors' Facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication & Society, 22*(10), 1497–1513. https://doi.org/10.1080/1369118X.2018.1445273

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

Hwang, H., Kim, Y., & Kim, Y. (2018). Influence of discussion incivility on deliberation: An examination of the mediating role of moral indignation. *Communication Research, 45*(2), 213–240. https://doi.org/10.1177/0093650215616861

Jin, S. V., Phua, J., & Lee, K. M. (2015). Telling stories about breastfeeding through Facebook: The impact of user-generated content (UGC) on pro-breastfeeding attitudes. *Computers in Human Behavior, 46*, 6–17. https://doi.org/10.1016/j.chb.2014.12.046

Jucks, R., Päuler, L., & Brummernhenrich, B. (2016). "I need to be explicit: You're wrong": Impact of face threats on social evaluations in online instructional communication. *Interacting with Computers, 28*(1), 73–84. https://doi.org/10.1093/iwc/iwu032

Kennedy-Lightsey, C. D. (2010). Recognizing contributions: Face-support and face-threat influences students' emotional and communicative responses. *Communication Research Reports, 27*(1), 20–29. https://doi.org/10.1080/08824090903526588

Kenski, K., Coe, K., & Rains, S. A. (2020). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research, 47*(6), 795–814. https://doi.org/10.1177/0093650217699933

Klofstad, C. A., Sokhey, A. E., & McClurg, S. D. (2013). Disagreeing about disagreement: How conflict in social networks affects political behavior. *American Journal of Political Science, 57*(1), 120–134. https://doi.org/10.1111/j.1540-5907.2012.00620.x

Lim, S. S., Vadrevu, S., Chan, Y. H., & Basnyat, I. (2012). Facework on Facebook: The online publicness of juvenile delinquents and youths-at-risk. *Journal of Broadcasting & Electronic Media, 56*(3), 346–361. https://doi.org/10.1080/08838151.2012.705198

Lin, C.-C., & Yamaguchi, S. (2011). Effects of face experience on emotions and self-esteem in Japanese culture. *European Journal of Social Psychology, 41*(4), 446–455. https://doi.org/10.1002/ejsp.817

Loveland, M. T., & Popescu, D. (2011). Democracy on the web. *Information, Communication & Society, 14*(5), 684–703. https://doi.org/10.1080/1369118X.2010.521844

Lu, Y., & Gall Myrick, J. (2016). Cross-cutting exposure on Facebook and political participation. *Journal of Media Psychology: Theories, Methods, and Applications, 28*(3), 100–110. https://doi.org/10.1027/1864-1105/a000203

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective intelligence and political judgement*. University of Chicago Press.

McDevitt, M., Kiousis, S., & Wahl-Jorgensen, K. (2003). Spiral of moderation: Opinion expression in computer-mediated discussion. *International Journal of Public Opinion Research, 15*(4), 454–470. https://doi.org/10.1093/ijpor/15.4.454

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior, 39*(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5

Mutz, D. C. (2002). The consequences of cross-cutting networks for political participation. *American Journal of Political Science, 46*(4), 838–855. https://doi.org/10.2307/3088437

Nabi, R. L. (2010). The case for emphasizing discrete emotions in communication research. *Communication Monographs, 77*(2), 153–159. https://doi.org/10.1080/03637751003790444

Neubaum, G., & Krämer, N. C. (2018). What do we fear? Expected sanctions for expressing minority opinions in offline and online communication. *Communication Research, 45*(2), 139–164. https://doi.org/10.1177/0093650215623837

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). https://doi.org/10.1111/j.1083-6101.2005.tb00252.x

Oetzel, J. G., Ting-Toomey, S., Masumoto, T., Yokochi, Y., Pan, X., Takai, J., & Wilcox, R. (2001). Face and facework in conflict: A cross-cultural comparison of China, Germany, Japan, and the United States. *Communication Monographs, 68*(3), 235–258. https://doi.org/10.1080/03637750128061

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

Orehek, E., Bessarabova, E., Chen, X., & Kruglanski, A. W. (2011). Positive affect as informational feedback in goal pursuit. *Motivation and Emotion, 35*(1), 44–51. https://doi.org/10.1007/s11031-010-9197-2

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

Price, V., Cappella, J. N., & Nir, L. (2002). Does disagreement contribute to more deliberative opinion? *Political Communication, 19*(1), 95–112. https://doi.org/10.1080/105846002317246506

Price, V., Nir, L., & Cappella, J. N. (2006). Normative and informational influences in online political discussions. *Communication Theory, 16*(1), 47–74. https://doi.org/10.1111/j.1468-2885. 2006.00005.x

Rains, S. A. (2013). The nature of psychological reactance revisited: A meta-analytic review. *Human Communication Research, 39*(1), 47–73. https://doi.org/10.1111/j.1468-2958.2012.01443.x

Reeves, B., & Nass, C. (1996). *The media equation*. Cambridge University Press.

Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review, 1*(1), 60–85. https://doi.org/10.1177/1754073908097189

Smith, C. B., McLaughlin, M. L., & Osborne, K. K. (1998). From terminal ineptitude to virtual sociopathy: How conduct is regulated on Usenet. In F. Sudweeks, M. L. McLaughlin, & S. Rafaeli (Eds.), *Networs and netplay: Virtual groups on the Internet* (pp. 95–112). MIT.

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

Stromer-Galley, J., & Muhlberger, P. (2009). Agreement and disagreement in group deliberation: Effects on deliberation satisfaction, future engagement, and decision legitimacy. *Political Communication, 26*(2), 173–192. https://doi.org/10.1080/10584600902850775

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

Valentino, N. A., Brader, T., Groenendyk, E. W., Gregorowicz, K., & Hutchings, V. L. (2011). Election night's alright for fighting: The role of emotions in political participation. *The Journal of Politics, 73*(1), 156–170. https://doi.org/10.1017/S0022381610000939

Vangelisti, A. L. (1994). Messages that hurt. In W. R. Cupach & B. H. Spitzberg (Eds.), *The dark side of interpersonal communication* (pp. 53–82). Lawrence Erlbaum.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Willer, E. K., & Soliz, J. (2010). Face needs, intragroup status, and women's reactions to socially aggressive face threats. *Personal Relationships, 17*(4), 557–571. https://doi.org/10.1111/j.1475-6811.2010.01297.x

Wojcieszak, M. E., & Price, V. (2012). Perceived versus actual disagreement: Which influences deliberative experiences? *Journal of Communication, 62*(3), 418–436. https://doi.org/10.1111/j.1460-2466.2012.01645.x

Zadro, L., Williams, K. D., & Richardson, R. (2004). How low can you go? Ostracism by a computer is sufficient to lower self-reported levels of belonging, control, self-esteem, and meaningful existence. *Journal of Experimental Social Psychology, 40*(4), 560–567. https://doi.org/10.1016/j.jesp.2003.11.006

Zhang, S., & Stafford, L. (2008). Perceived face threat of honest but hurtful evaluative messages in romantic relationships. *Western Journal of Communication, 72*(1), 19–39. https://doi.org/10.1080/10570310701828628

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]. Springer.

Ziegele, M., Breiner, T., & Quiring, O. (2014). What creates interactivity in online news discussions? An exploratory analysis of discussion factors in user comments on news items. *Journal of Communication, 64*(6), 1111–1138. https://doi.org/10.1111/jcom.12123

Ziegele, M., & Jost, P. (2020). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research, 47*(6), 891–920. https://doi.org/10.1177/0093650216671854

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

Ziegele, M., Weber, M., Quiring, O., & Breiner, T. (2018). The dynamics of online news discussions: Effects of news articles and reader comments on users' involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society, 21*(10), 1419–1435. https://doi.org/10.1080/1369118X.2017.1324505

History
Received June 6, 2020
Revision received July 6, 2021
Accepted October 31, 2021
Published online February 2, 2022

ORCID
Teresa K. Naab
https://orcid.org/0000-0001-7345-2559

Teresa K. Naab
Department of Media, Knowledge and Communication
University of Augsburg
Universitätsstrasse 10
86159 Augsburg
Germany
teresa.naab@phil.uni-augsburg.de

Teresa K. Naab (PhD) is a researcher at the Department of Media, Knowledge and Communication at the University of Augsburg, Germany. She holds a PhD in Communication Science from Hanover University of Music, Drama and Media, Germany. Her research focuses on digital communication and quantitative social science methods.