



How social are self-effects? The impact of feedback on the internalization of expressed opinions in online communication

Stephan Winter^{a,*}, Anne L. Vos^b, Paola Remmelswaal^b, Peter Neijens^b

^a University of Kaiserslautern-Landau (RPTU), Department of Psychology, Germany

^b University of Amsterdam, Amsterdam School of Communication Research, Netherlands

ARTICLE INFO

Keywords:

Opinion formation
User comments
Self-effects
Social media
Persuasion
Experiment

ABSTRACT

Online communication channels provide rich opportunities for public opinion expression, and the act of posting may lead to an internalization of the expressed views. This research investigated whether such self-effects are affected by feedback comments and whether this influence is stronger in social network sites than in anonymous online forums. In an experiment manipulating valence of feedback and the media context, 161 participants were asked to express their opinion on a current topic. Results showed limited effects of feedback: Negative comments tended to reduce internalization among participants with non-extreme prior attitudes, but this was independent from the media context.

One central characteristic of online communication is the opportunity to post content to a potentially large audience: Compared to traditional mass media, the barriers for content production have been drastically reduced and it is relatively easy to express one's opinion, for instance, on social network sites (SNS). While most studies focus on the consequences of exposure to content on these channels (e.g., whether people are persuaded by user-generated content even if it contains low-quality information), researchers have begun to also pay attention to the effects that posting something may have on the sender: How does the act of expressing an opinion affect oneself? This phenomenon has been termed self-effect (Valkenburg, 2017) or expression effect (Pingree, 2007).

Initial studies have shown that people tend to internalize behavior they showed in front of relevant others or self-descriptions they posted online: For instance, someone who had the task to describe her-/himself as outgoing (e.g., by writing about social activities) expressed higher extraversion self-assessments afterwards than someone who had the task to describe introverted facets of her/his personality (e.g., Carr & Foreman, 2016; Gonzales & Hancock, 2008). Most studies in this so-called identity shift paradigm focused on extra-/introversion but recent work indicated that the process may also occur for opinions toward current news topics and public debates (Cho, Ahmed, Keum, Choi, & Lee, 2018; Nekmat, 2012; Winter, Remmelswaal, & Vos, 2022). However, the factors that affect the magnitude of such self-effects are yet to be explored.

First, it is likely that online opinion expressions do not stand alone but receive feedback (Brady, McLoughlin, Doan, & Crockett, 2021; Walther et al., 2011), for instance, in the form of likes, shares, or (positive and negative) comments. Even if the content of the opinion expression did not mirror one's private attitude in the first place (but was affected by social goals), receiving positive feedback may amplify an internalization of the expressed opinion. On the contrary, self-effects may be diminished by negative comments – at least if the originator cares about the social group and does not feel strongly about the topic. Our first goal is therefore to study the impact of feedback valence on self-effects.

Second, effects may depend on the media context in which opinions are expressed and feedback is received. On SNS, users are typically connected to a profile with their real name or a nickname that can be recognized by their friends and acquaintances. SNS as a popular subcategory of social media (Bayer, Trieu, & Ellison, 2020; Carr & Hayes, 2015) can be differentiated from more classic online forums, which offer a similar functionality of posting and commenting but allow for anonymity. Thus, SNS users experience higher levels of identifiability and connections to a relevant audience (Bayer et al., 2020) and therefore probably also perceive the feedback received in that community as important. A potential consequence could be that positive feedback in predominantly like-minded circles may strengthen or polarize existing attitudes. On the other hand, counter-speech or opposing comments from friends or other peers may also lead people to reconsider the

* Corresponding author. University of Kaiserslautern-Landau (RPTU), Department of Psychology, Fortstr. 7, 76829, Landau, Pfalz, Germany.

E-mail address: stephan.winter@rptu.de (S. Winter).

<https://doi.org/10.1016/j.chbr.2023.100365>

Received 15 July 2022; Received in revised form 21 December 2023; Accepted 22 December 2023

Available online 23 December 2023

2451-9588/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

original ideas. Our second goal is therefore to study the impact of the media context (SNS vs. forum) on the occurrence of feedback effects.

In the following, we will discuss theoretical explanations for self-effects and the role of feedback in identifiable vs. anonymous settings. The derived predictions will be tested with a laboratory experiment in which participants are asked to express their opinion on a news topic either in an SNS group or an online forum and receive positive, negative or no feedback on their post. Based on the results, we will discuss implications for the analysis of self-effects and the dynamics of opinion formation in social media.

1. Self-effects and internalization of expressed opinions

Although communication research typically focuses on the question of how messages affect their *recipients*, some studies and recent theorizing also paid attention to effects that messages have on their *senders*. This research has become increasingly relevant in times of online communication and social media, as contemporary media environments foster user activities and the production of user-generated content. The question of how creating messages affects their senders (e.g., by making them more extreme due to internalization) thus contributes to a more comprehensive understanding of digital media effects, since a focus on reception effects would neglect core characteristics of today's media usage.

According to Valkenburg (2017), a self-effect is given when "message creators/senders involuntarily influence their own cognitions, emotions, attitudes, or behavior" (p. 478). Pingree (2007) differentiated typical phases of sending messages: 1) expecting an expression, 2) composing a message, and 3) releasing it. Self-effects may already occur in the first two stages (e.g., an imagined complaint to one's boss or mere diary writing may reduce stress) but are likely most potent when the message is actually received by others (expression effect; Pingree, 2007). The pattern that behavior shown in public shapes people's subsequent self-perceptions was already documented in classic social psychological studies (Baumeister & Tice, 1984; Tice, 1992), for instance, that those who had the task to describe themselves in a certain manner (not by lying but by highlighting one's extraverted [vs. introverted] sides, for example) also evaluated themselves in line with this task (i.e., they reported higher extraversion levels in a personality questionnaire). Tice (1992) argued that such an internalization of behavior or self-descriptions is more likely when the content of the self-presentation is not made-up but the person has high degrees of freedom (and provides selective but authentic information), when this self-presentation can be connected to one's identity, and when it is seen by a relevant audience.

In online communication, such messages about one's personality or one's views can easily be posted, and the opportunities for selective self-presentation (Walther, 1996) fulfil the condition of high degrees of choice named above. Furthermore, it can be argued that online messages are typically more persistent than in face-to-face communication. While something said in a short face-to-face encounter with friends can easily be neglected, posts on SNS are visible for longer, and the creator of a message may also be confronted with it at a later occasion (Carr, 2021). For these reasons, it seems plausible that such self-effects are also (or even more) likely in online communication. The pioneer study by Gonzales and Hancock (2008) showed that induced behavior (providing a self-description as outgoing vs. shy, as in the "offline" studies) led to corresponding changes in self-perceptions when the message was posted on a public blog. The identity shift did not occur when the message was only posted in a private text document. Subsequent studies showed similar identity shifts induced by SNS posts (e.g., Carr & Foreman, 2016) and also showed self-effects for other (non-personality) outcomes such as health behavior (Nabi, Huskey, Nicholls, Keblusek, & Reed, 2019), relationship longevity (Toma & Choi, 2015) and attitudes toward news topics or politics (Cho et al., 2018; Lane et al., 2019; Winter, Remmelswaal, & Vos, 2022). Walther et al. (2018) argued that observing oneself produce a message in computer-mediated

communication, "for whatever reason it was produced, may lead the actor to infer that he or she feels whatever it is that may be reflected in that message" (p. 309).

Self-effects have generally been explained by self-perception theory (Bem, 1972): In this realm, individuals are assumed to be observers of their own behavior who infer their traits based on this observation (and recent public behavior is assumed to be most salient). However, the pattern that effects are smaller or absent when no one is watching/reading (Gonzales & Hancock, 2008; Tice, 1992) cannot fully be explained by this approach and thus gives credence to other explanations such as theoretical accounts of public commitment (Schlenker, Dlugolecki, & Doherty, 1994), assuming that people feel bound to statements they made in public (to avoid being perceived as inconsistent) and strive for self-perceptions that are in line with their presumed public image.

In the following, we focus on the internalization of attitudes after opinion expression as a self-effect that is particularly relevant in the realm of discussions on news or public affairs topics in social media (Lee & Shin, 2021). Internalization would be given if the sender's private attitude toward a topic becomes more positive (negative) after releasing a positive (negative) message about the issue. For example, when a Facebook user is a moderate proponent of stricter environmental restrictions and posts a positive statement about a new law (perhaps to please some friends from their contact list that are environmental activists), he/she may become a stronger proponent of the law. Lee and Shin (2021) aptly described this process as "I share, therefore, I believe (more)" (p. 268).

1.1. The impact of feedback

So far, self-effects can be regarded as mainly *intrapersonal*, although they are always assumed to be social in the sense that the existence of an (imagined) audience amplifies them. However, when considering the typical usage situation, it is common that opinion expressions receive feedback by others in the form of comments or likes, which makes it an explicitly *interpersonal* process. The hyperpersonal model (Walther, 1996) highlights feedback as a central element of computer-mediated communication (besides sender, receiver, and channel) and holds that it can reinforce the selective self-presentation given by the sender. Along these lines, Walther et al. (2011) extended the study by Gonzales and Hancock (2008) and found that positive feedback (in this case: confirming that the originator of the blog post is extraverted/introverted) leads to stronger identity shifts. Johnson and van der Heide (2015) investigated the effect of feedback on so-called taste performances: Participants selected an artistic photograph and then received ostensible feedback about their choice. When they received positive feedback, participants tended to like the photo more than the non-selected photos in a later questionnaire, and negative feedback directly led to lower liking.

Transferring these assumptions to the scenario of discussing news topics online, we expect that feedback also affects the internalization of expressed opinions. When a post about a topic is commented positively, the sender may feel confirmed that this is a socially desirable point of view, which may also help to create a positive impression of the sender. Referring to the theoretical explanations for self-effects mentioned above, positive feedback likely leads to higher attention for the specific statement (and makes it more salient in self-perception) and also increases public commitment (contradicting a public statement that has been praised by others would seem particularly inconsistent). A recent content analysis of moral outrages on Twitter (Brady et al., 2021) demonstrated that users who received many likes for outrageous statements were more likely to show such posting behavior in the following days. However, this does not necessarily mean that users' private attitudes also became more extreme. We therefore aim to test the effect of feedback on the internalization of an expressed opinion and posit the following hypothesis concerning the resulting privately held attitude

and the confidence with which it is held (Lee & Shin, 2021). Positive feedback is expected to amplify the presumed self-effect of internalization and additionally increase attitude certainty (the degree to which people are convinced that their attitude toward the topic is correct), which has been found to depend on the degree of approval of this attitude in a social context (Ryffel, Wirz, Kühne, & Wirth, 2014).

H1. Positive feedback leads to a stronger internalization of an expressed opinion than negative or no feedback.

H2. Positive feedback leads to higher attitude certainty than negative or no feedback.

1.2. Feedback in SNS and forums: The role of anonymity

The public commitment approach (Schlenker et al., 1984) holds that self-effects occur more frequently after public rather than private behavior, and publicness is also a prerequisite for feedback to occur (Walther et al., 2011). What has remained unclear so far is whether the degree of publicness – or the media context in which opinion expressions take place – affect the importance of feedback in the process of internalizing an opinion. We focus on SNS, which can be defined as web-based services in which users create (semi-)public profiles, share content, and are connected with other users (Boyd & Ellison, 2007), as a very popular form of contemporary social media formats. As a comparison category, we refer to online forums as an earlier format of computer-mediated communication that similarly allows for sharing content and posting comments but typically works with anonymous nicknames. SNS users are more identifiable, and thus, there is a higher importance of social relationships and self-presentation (Nadkarni & Hofmann, 2012), and one's own behavior is visible to relevant others (Leonardi & Treem, 2012).

On this basis, it is plausible to assume that the social cues that are salient in SNS environments also lend more importance to the feedback given by other users. While comments by anonymous forum users can be discarded quite easily, SNS commenters typically appear to be more relevant or relationally closer (Carr & Foreman, 2016) and the feedback relates to an opinion expression that is directly tied to one's profile and thus to one's identity (see Tice, 1992). Therefore, we propose that:

H3. The effect of feedback on users' internalization of the expressed opinion is stronger in (non-anonymous) SNS than in an anonymous forum.

2. Method

2.1. Participants and design

A laboratory experiment in which participants were asked to express their opinion on a current topic was conducted to test the proposed hypotheses. In a 3x2 between-subjects design, the valence of feedback given to the opinion expression (positive vs. negative vs. no feedback) and the media context (SNS vs. online discussion forum) were systematically varied. Participants were randomly assigned to one of the six conditions.

Students appeared to be an appropriate target group, since SNS and online forums are popular and common news sources among the younger population (Mitchell, Gottfried, & Matsa, 2015). Therefore, we recruited participants in Facebook groups of the university and on campus. The requirements for participations were the status of a Bachelor student at the University of Amsterdam and the possession of a personal Facebook profile, which led to convenience sample of this group of students. Participants either received 10 Euro or course credits as compensation. The study procedure was approved by the local ethics committee.

In total, 161 students participated in the experiment (121 female; age: $M = 21.38$, $SD = 2.37$; period of data collection: April–September

2018). With regard to the usage of SNS and the importance of Facebook, participants reported a moderate intensity ($M = 2.66$, $SD = 0.96$, for the frequency of news sharing and commenting on SNS, measured by six items such as “How often do you comment on others' postings on Facebook?”, and a mean value of 3.50 ($SD = 1.29$) on the Facebook Intensity scale with six items such as “I feel uninformed if I have not logged in to Facebook for a long time” (Ellison, Steinfield, & Lampe, 2007; $\alpha = 0.83$)).

2.2. Topic and stimulus material

The experiment was presented to participants as a study on an online discussion about current debates. The topic for the online discussion was derived from a prior study (Winter, Remmelswaal, & Vos, 2022): a proposed ban of bicycle-sharing services in Amsterdam. A pilot study had shown that the topic is interesting for the target group and students indicated moderate levels of prior attitudes. In the beginning, participants were asked to read two articles about the bicycle-sharing issue to get an overview of the topic. The texts were shown in the design of news articles and included arguments in favor (such as “shared bikes sometimes lie in public parking spots for days, leading to complaints by residents”) and against the bicycle-sharing ban (such as “shared bikes provide a convenient transportation mode for both citizens' and tourists' commutes”).

2.3. Independent variable: Media context

The online discussion took place in either an SNS group or an online discussion forum. The SNS context was represented by a Facebook group with Bachelor students of the University of Amsterdam. The group was created for this study and filled with more than 250 profiles of Facebook users to create a realistic usage scenario. In the experiment, participants were asked to log in to their personal Facebook accounts and join the Facebook group. In the group, participants were asked to answer a question posted by one of the users (an ostensible student account that was designed by the experimenters to look as realistic as possible, including pictures and background information): “I read Amsterdam plans to ban bicycle-sharing services from the city. What are your opinions on that? Are you in favor or against the ban?” (see Fig. 1).

In the online discussion forum, named “The online student board”, described as a general platform on which students can anonymously talk about university issues but also about other current debates, participants were asked to answer the same question. The key difference to the SNS condition was that the question was written by a post initiator with a nickname that did not reveal personal information. Similarly, participants in the forum condition were instructed to answer to the post with an anonymous nickname of their choice (see Fig. 2).

2.4. Independent variable: Valence of feedback

After participants wrote down their opinion, they received either positive, negative or no feedback. In the positive feedback condition, participants received the following comment, which was dynamically adapted by the experimenter based on the valence of their expressed opinion: “Very convincing post! I totally agree with your view! I think there are too many disadvantages/still many advantages of the shared bicycles and it would be logical/illogical to ban them. It's very annoying that they take so much space ... /They are convenient and help the environment ...”, whereas in the negative feedback condition, participants received the following tailored feedback: “That doesn't convince me, I don't agree at all! I think there are too many disadvantages/still many advantages of the shared bicycles and it would be logical/illogical to ban them. It's very annoying that they take so much space ... /They are convenient and help the environment ...”. In addition, on Facebook, participants in the positive feedback condition received a like for their comment.

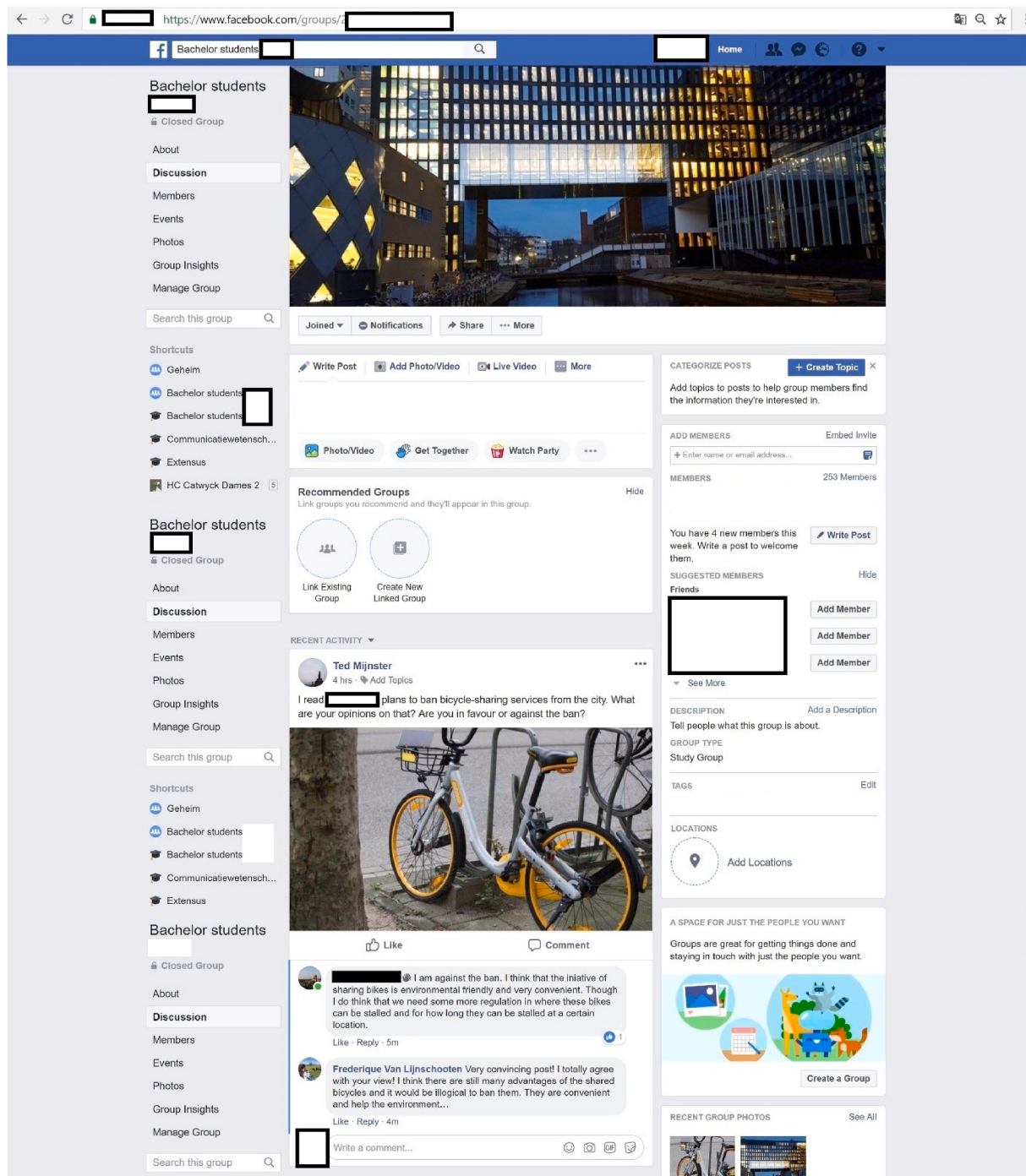


Fig. 1. The SNS context: a Facebook group for Bachelor students of the university (names and location blinded).

The feedback comment and like in the Facebook condition was given by an ostensible student profile (the profile had a significant number of friends, a personal biography, and multiple photos to appear realistic). In the forum condition, participants received feedback from a user with an anonymous nickname “Student22”.

2.5. Procedure

Participants first read two news articles on the topic. After that, the questionnaire started with questions about their attitude toward the bicycle-sharing issue. Then, participants were directed to either the Facebook group or the online forum and asked to write down their opinion on the bicycle-sharing ban. They were instructed to take a clear

position in favor or against the ban. Then, participants continued the questionnaire that contained some distraction questions about the text quality and participants’ need for cognition. This way, the lab assistant had enough time to read the expressed opinion and provide tailored feedback accordingly. Once participants finished answering the distraction questions, they were redirected to the Facebook group or online forum to read the feedback that they had been given. Next, participants were asked to continue with some filler questions measuring their attitude towards other debatable topics (such as health data and privacy), which were followed by questions measuring participants’ attitude towards the bicycle-sharing ban for a second time and their attitude certainty. Finally, participants answered questions about manipulation checks, personality characteristics, connection to the topic

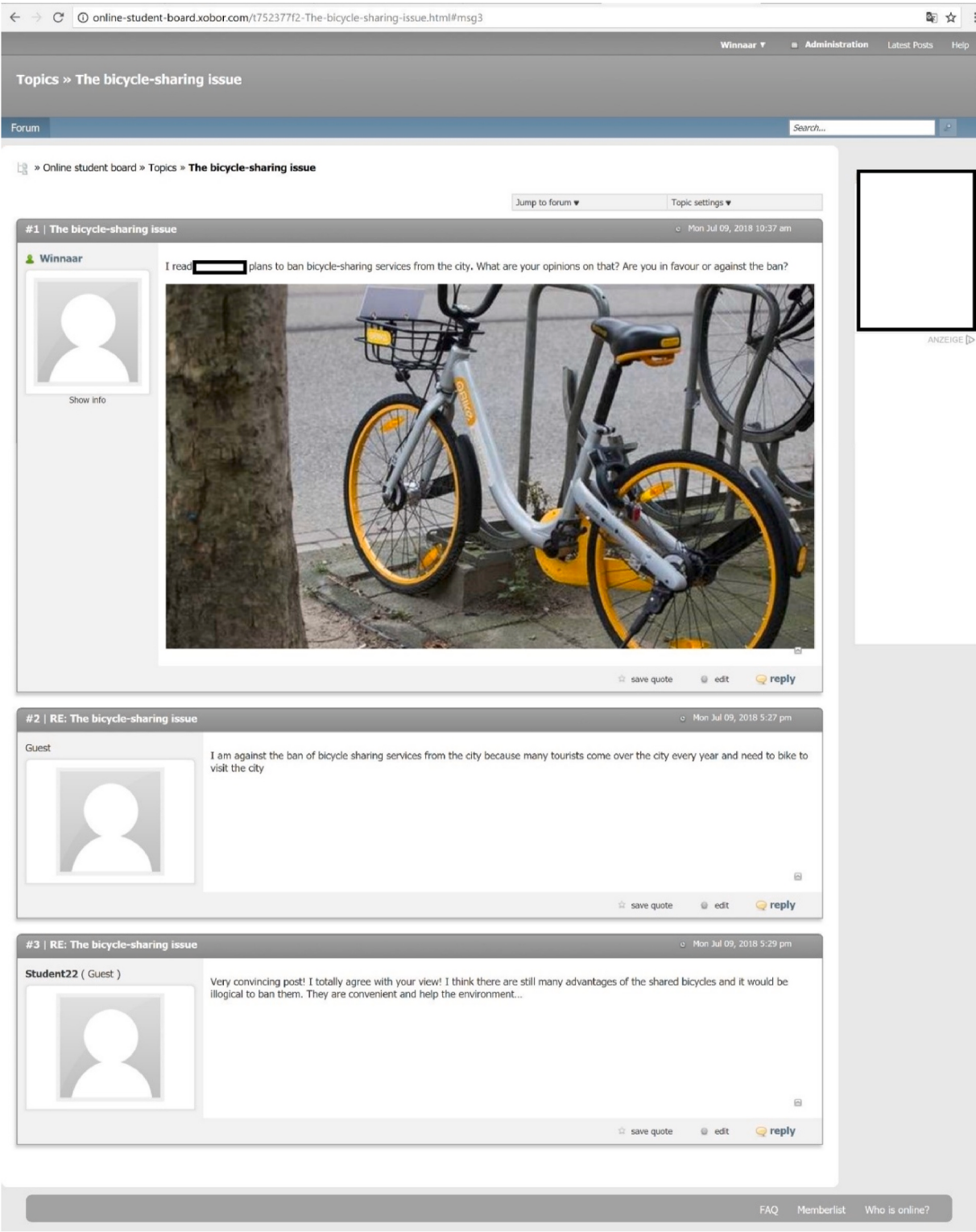


Fig. 2. The forum context: an online discussion website.

(e.g., prior usage of bicycle-sharing services) and further demographics.

2.6. Measures

Valence of the opinion. The valence of the expressed opinion was coded by two research assistants. Based on the scheme by Winter, Remmelswaal, & Vos, (2022), the overall opinion was coded as either negative (e.g., negative comments about the ban, positive comments

about the shared-bicycle system, $n = 102$), neutral (equal amount of pros and con, $n = 0$), or positive (positive comments about the ban, negative comments about shared bicycles, $n = 59$). After coding approximately 10% of the final sample, the intercoder reliability test showed 100% agreement.

Attitude toward the topic. Participants' attitude toward the bicycle-sharing ban was measured with three items ("Bicycle-sharing services in Amsterdam should be banned", "Bicycle-sharing services in Amsterdam

are unacceptable”, “The proposition to ban bicycle-sharing services in Amsterdam should be supported”, rated on a scale from 1 (strongly disagree) to 7 (strongly agree; Winter, Remmelswaal, & Vos, 2022). The items were measured before and after opinion expression (Cronbach's α at $T_1 = 0.90$, $M = 3.17$, $SD = 1.44$; α at $T_2 = 0.90$, $M = 3.25$, $SD = 1.59$).

Attitude certainty. One item measured participants' attitude certainty (“How certain are you about your own opinion regarding the bicycle-sharing issue?”), which was rated on a scale from 1 (very uncertain) to 7 (very certain) ($M = 5.46$, $SD = 1.18$).

Internalization. The calculation of the internalization score included three steps. First, we aimed to filter out participants with an overall neutral statement (as such a statement cannot induce opinion shifts in a positive or negative direction/however, neutral statements did not occur in this sample) as well as participants with a contradictory attitude score (e.g., their attitude toward the topic at T_1 was negative, but their expressed opinion was positive toward the ban, or vice versa/for these participants, the classification as positive/negative feedback regarding their own attitude may not be correct; $n = 11$). Second, the attitude score at T_1 was subtracted from the attitude score at T_2 : If participants had a higher attitude after opinion expression with a positive post, this indicates that they internalized the expressed opinion and moved further in this direction. Third, for participants who expressed a negative opinion toward the ban, the resulting difference score was multiplied by -1 , to guarantee that higher values indicate a shift of participants' attitude in the direction of the statement ($M = 0.21$, $SD = 0.73$, $\min = -1.33$, $\max = 2.00$).

2.7. User characteristics

Self-monitoring. The self-monitoring scale, measuring participants' propensity to adapt their behavior to the environment, consisted of 18 items (e.g., “I can fool people by showing myself friendly, even though I really do not like them”, 7-point scale; Snyder, 1974; $\alpha = 0.73$, $M = 4.15$, $SD = 0.70$).

Fear of negative evaluation. People's anxiety to be judged negatively in social situation was measured with the Fear of Negative Evaluation scale (Leary, 1983; 12 items, e.g., “I am frequently afraid of other people noticing my shortcomings”, 7-point scale, $\alpha = 0.94$, $M = 3.99$, $SD = 1.28$).

2.8. Manipulation checks

Valence of the articles. The perceived valence of the articles was assessed with one item (“Overall, the articles were ... about the proposal to ban bicycle-sharing services”, rated on a scale from 1 (extremely negative) to 7 (extremely positive), $M = 4.17$, $SD = 1.25$).

Valence of the feedback. One item measured participants' assessment of the valence of the feedback (“Overall, the feedback given by other students (on the comment you posted) was ...”, rated on a scale from 1 (extremely negative) to 7 (extremely positive), $M = 4.50$, $SD = 2.09$).

Identification with group members. The identification with other members in the group was assessed with six items (e.g., “I feel a bond with these people”, 7-point scale, Doosje, Ellemers, & Spears, 1995; Lee, 2004, α for Facebook members = 0.87, α for forum members = 0.87, combined scale: $M = 3.66$, $SD = 0.94$).

Public identifiability. The extent to which participants believed that their comment could be linked to their name was measured with one item (“How likely do you think it is that your own written comment can be linked to your name?”), rated on a scale from 1 (extremely unlikely) to 7 (extremely likely), based on Walther et al., 2011; $M = 4.44$, $SD = 2.06$).

3. Results

3.1. Randomization and manipulation checks

Analyses showed no significant differences between the experimental conditions in gender, $\chi^2(5) = 3.59$, $p = .610$, age, $F(5, 155) = 0.81$, $p = .541$, and prior usage of the bicycle-sharing services $\chi^2(5) = 2.80$, $p = .731$. This indicates a successful randomization.

Several variables served as manipulation checks. First, an analysis of variance (ANOVA) with the experimental condition (combinations of media context and valence of feedback) as independent variable and the perceived valence of the articles as dependent variable showed no significant effect, $F(5, 155) = 1.17$, $p = .328$, $\eta_p^2 = 0.04$, indicating that the articles were perceived as equally neutral ($M = 4.17$, $SD = 1.25$) across conditions. Second, a t -test with the manipulated valence of feedback as independent variable and the perceived valence of the feedback as dependent variable showed that the experimental groups differed significantly from each other, $t(106) = -19.16$, $p < .001$: As intended, perceptions of feedback in the negative condition ($M = 2.51$, $SD = 1.19$) were more negative than in the positive condition ($M = 6.60$, $SD = 1.03$). Third, perceived identifiability of one's post differed across the two media groups, $t(159) = 7.25$, $p < .001$. More precisely, participants in the SNS condition believed that their comment was more publicly identifiable ($M = 5.46$, $SD = 1.72$) than participants in the forum condition ($M = 3.41$, $SD = 1.86$). This indicates a successful manipulation. We additionally tested whether the degree of identification with the other members of the group depended on the valence of feedback and the media context. Results showed a significant effect of valence, $F(2, 155) = 5.89$, $p = .003$, $\eta_p^2 = 0.07$: Participants expressed lower identification when negative feedback was given ($M = 3.31$, $SD = 1.05$; post-hoc comparisons with Bonferroni correction $p < .05$) than when positive ($M = 3.80$, $SD = 0.97$) or no feedback ($M = 3.86$, $SD = 0.66$) was given.

3.2. Internalization and attitude certainty effects

The first hypothesis stated that positive feedback leads to a stronger internalization of the expressed opinion than negative or no feedback. According to the analysis plan, the presumed effect was tested with an ANOVA with the valence of feedback as independent variable and internalization as dependent variable. Results showed no significant effect, $F(2, 147) = 2.12$, $p = .124$, $\eta_p^2 = 0.03$. Only in a second analysis, in which people with extreme attitudes at T_1 (1 or 7, $n = 10$) were excluded (as their attitudes cannot move further into the respective direction; Winter, Remmelswaal, & Vos, 2022), the effect of feedback was significant, $F(2, 137) = 3.42$, $p = .036$, $\eta_p^2 = 0.05$. According to the mean values, internalization was higher with positive ($M = 0.29$, $SD = 0.62$) and no feedback ($M = 0.31$, $SD = 0.58$) than with negative feedback ($M = 0.02$, $SD = 0.56$; post-hoc comparisons with Bonferroni correction $p = .073$ for negative vs. no feedback and $p = .079$ for negative vs. positive feedback). The pattern for positive and negative feedback was in line with predictions, but the small difference and the overall pattern only provide limited support H1.

The second hypothesis posited that positive feedback leads to higher attitude certainty than negative or no feedback. To test the presumed effect, an ANOVA with feedback valence as independent variable and attitude certainty as dependent variable was conducted. However, against our expectations, the results showed no significant effect, $F(2, 158) = 0.29$, $p = .750$, $\eta_p^2 = 0.004$, meaning that H2 needs to be rejected.

In the third hypothesis, we assumed that the effect of feedback on users' internalization of the expressed opinion would be stronger in SNS than in an anonymous forum. To test this specific interaction pattern, a planned contrast analysis with the expected directional effects was constructed. Six contrast weights reflect the conditions of the 3x2 design. The specific weights for each contrast were calculated in two steps. First, weights were derived for the effect of each factor (i.e., the

media context and valence of feedback) in each condition. Second, the preliminary weights of all these factors were summed in each cell (Rosenthal & Rosnow, 1985, see Table 1). Hence, a preliminary weight of 1 was given to conditions with positive feedback, a weight of -1 to conditions with negative feedback and a 0 for conditions with no feedback. Next, since it was expected that the presumed effect of feedback valence would be stronger in SNS than in forums, an additional weight of $+1$ (positive feedback) and -1 (negative feedback) was assigned to the SNS conditions. Finally, the weights of all factors were summed and tested in a contrast analysis. The result of the contrast analysis was only marginally significant, $t(144) = 1.83$; $p = .070$ and $t(134) = 1.95$; $p = .053$ (for the analysis without participants whose initial attitudes could not become more extreme), and mean values (see Table 1) did not show the pattern that the effect of feedback (e.g., the difference between positive and negative feedback) was larger in the SNS conditions. Therefore, H3 is not supported by the data. The original hypothesis only referred to internalization as dependent variable and not to attitude certainty (as the latter was assumed to be a more heuristic evaluation based on the valence of feedback). However, we also explored whether feedback in SNS (vs. forums) has a stronger effect on one's attitude certainty (following the same logic of H3). The contrast analysis did not yield a significant result, $t(155) = 0.41$; $p = .683$.

To additionally test whether the effect of SNS feedback on users' internalization of the expressed opinion is stronger for people with a) higher self-monitoring and b) higher fear of negative evaluation, moderation analyses using the PROCESS macro were conducted. Valence of feedback (dummy-coded) was treated as independent variable, and media context as well as self-monitoring/fear of negative evaluation as moderators; the dependent variable was internalization. However, for both potential moderators, no significant interactions emerged.

4. Discussion

The goal of the present study was to investigate the impact of feedback (valence and media context) on the internalization of expressed opinions in online communication channels. In an experiment, participants expressed an opinion about a current exemplary topic (a proposed ban of bicycle-sharing systems), either in an SNS group or an anonymous online forum, and then received positive, negative or no feedback on their post. Given the interactivity of online communication, the study sought to analyze how self-effects (predominantly characterized as an intrapersonal process) depend on social reactions in the form of comment and likes – this also refers to the practically relevant question of how confirmatory comments that users receive on their opinion expression may lead to further shifts in attitudes and increases in polarization.

The concept of self-effects recently received scholarly attention (e.g., Valkenburg, 2017) as public expressions are particularly frequent in social media and perhaps even more important for the analysis of opinion formation in contemporary media environments than classic reception effects. Recent studies indicate that the act of posting something online can indeed affect private attitudes by shifting them in the direction of the posted opinion (Cho et al., 2018; Winter, Rummelsaal,

& Vos, 2022) – extending prior findings on effects of (induced) self-presentational behavior on one's self-concept (Walther et al., 2011), this means that self-effects are also relevant for the formation of views on news topics or political issues (Lane et al., 2019). The positive mean value of the internalization index in the sample of this study also suggests a (small) shift of attitudes across conditions.

With regard to the importance of feedback to the opinion expression, results only showed a limited effect of social reactions: The effect of the manipulation was only significant in the subsample of participants whose prior attitudes were not already at the extreme sides of the spectrum. Among these participants, the extent of internalization was higher after receiving a positive comment or no comment at all than after receiving a negative comment. We had expected that positive feedback would strengthen internalization, as a confirming comment should reassure the social desirability of the public image and lead to higher attention to the statement (and make it more salient in self-perception) as well as to higher public commitment (as contradicting a public statement that has been praised by others would seem particularly inconsistent; Lane et al., 2019; Schlenker et al., 1984; Tice, 1992). However, the mean values did not differ between positive and no feedback, and if at all, the pattern suggests that negative feedback led to lower internalization in comparison to the other conditions. One tentative conclusion could be that self-effects are not reinforced by positive comments (the act of posting something without any feedback may already be sufficient) but that negative comments may diminish this process. This would suggest that self-effects can indeed also be interpersonal (and not only intrapersonal), which is in line with the hyperpersonal model (Walther, 1996) that highlights feedback as a central component of computer-mediated communication. However, it has to be noted that the effect was small and only occurred in a subsample.

Previous studies in the realm of identity shifts (self-presentations as extravert/introvert: Carr & Foreman, 2016; Walther et al., 2011) as well as in the scenario of sharing media content (selecting abstract photographs: Johnson & van der Heide, 2015) showed stronger effects of peer feedback. This discrepancy may be explained by the domain: Feedback is likely to be more important for self-assessments concerning social behavior or taste performances, whereas social approval may be slightly less relevant for the evaluation of a news topic (particularly after having read news articles on the topic) because attitudes on such current debates may depend more on the available arguments or general ideological worldviews. Against expectations, positive feedback did not increase attitude certainty. Although this meta-cognition typically depends on the degree of perceived social approval (Ryffel et al., 2014), it was not significantly affected by the received feedback. Perhaps one comment was not sufficient to signal such social approval, or the topic was not consequential enough to arouse uncertainty in the first place (the mean value of attitude certainty across conditions was 5.46 on a 7-point scale).

Due to the higher identifiability of users in SNS, it was assumed that the valence of feedback is more important in this setting than in an anonymous forum. However, although opinion expression in SNS groups (and thus also the feedback) is more strongly connected to one's identity and the audience is more relevant (Carr & Foreman, 2016; Tice, 1992), results did not show meaningful differences in the impact of feedback in SNS vs. forum settings, which suggests a more general process.

These findings may mitigate assumptions that social media particularly amplify normatively undesirable patterns of strengthening prior attitudes in homogeneous communication environments. We do not find evidence that self-effects and the impact of (supportive) feedback by others are stronger in SNS vs. in more traditional online forums. Instead, the internalization of attitudes may occur across contexts, as long it is not hampered by explicitly negative comments of others. The normative implications of this pattern for deliberative democracy mainly depend on the composition of the group in terms of ideological homo- or heterogeneity (Neubaum & Krämer, 2017): While internalization probably

Table 1

Weights and descriptive statistics for effects of media context and valence of feedback on internalization ($n = 140$).

Media context	SNS group			Forum		
	Pos	Neg	No	Pos	Neg	No
Feedback	1	-1	0	1	-1	0
Feedback X media context	1	-1	0	0	0	0
Contrast total	2	-2	0	1	-1	0
Internalization (M)	0.24	0.04	0.35	0.35	0.00	0.27
Internalization (SD)	0.67	0.65	0.61	0.57	0.48	0.57

occurs more frequently in like-minded groups with positive feedback, such self-effects are less likely in mixed groups in which commenters also contradict one's point of view.

4.1. Limitations

It has to be noted that the experiment, similarly as previous studies on self-effects, only covered a short period of interaction (posting one statement and receiving one comment) to isolate causes and effects – this does not preclude that repeated feedback exerts stronger effects. Although the study was conducted in a natural setting and participants used their actual Facebook accounts, comments were given by ostensible other group members and not by personally known friends. Thus, although there were no differences between SNS and forums in an experimental setting that presented identical content and feedback across media contexts, it is conceivable that the possibility to connect with one's friends (and the algorithmic curation) leads to different content or that the affordances of SNS make opinion expression and thus the observed internalization of attitudes more likely in the wild. Therefore, it would be a valuable next step to integrate assumptions of reception effects and self-effects (Valkenburg, 2017). For example, do social media platforms affect the perceived opinion climate (in relevant sub-groups) and how do the resulting perceptions of being in the majority/minority affect sharing of opinions? Which feedback do people predominantly receive and how does it reinforce self-effects? And how does positive feedback also lead to more openness for reception effects with regard to postings by like-minded others? A methodological challenge is to disentangle selection, reception and self-effects. In the present study, we aimed to address this issue by measuring the difference between reported attitudes before posting and after message release and reception of feedback – however, it is conceivable that already the first attitude measurement was affected by the anticipation of interaction (Pingree, 2007). This potential intricacy becomes even more complex outside the laboratory situation and calls for triangulation with qualitative research (such as interviews and think aloud data), survey assessments, behavioral data (analyses of posting behavior), and further experimental settings.

Apart from the focus on a short part of the process, the present study is limited by the use of only one exemplary topic and the small sample size (originating from a specific student group) due to the relatively effortful lab procedure and the need for a shared group environment. The process of internalizing expressed opinions and corresponding feedback may also occur for more divisive issues such as climate change or vaccinations but probably requires repeated interactions with relevant others; also, backfire effects after negative feedback by non-like-minded others would be possible when prior attitudes are strong. In this study, all participants were asked to write down their opinion, and although there were no indications that participants felt discomfort with this task, probably only few of them would have chosen to post something outside the lab. Furthermore, as there was no condition without opinion expression, it cannot be clearly disentangled whether the posted comment worked as a reinforcement of the self-effect or rather as persuasion due to the arguments within the comment. We believe that such persuasive reception effects were unlikely in this experimental setting since it was a relatively generic comment but it cannot be ruled out strictly without a reception-only condition.

With regard to the manipulation of the media context (SNS vs. forum), it has to be noted that the mean value of perceived identifiability in the forum condition was still on a medium level. This could be due to the choice of nicknames that may still have some connection to their identity (although the instruction explicitly asked them to choose an anonymous nickname). Future research could employ procedures without any nicknames at all or check which cues about the users are revealed in their nicknames.

5. Conclusion

With these limitations in mind, the present study provides insights on the social nature of self-effects and provides avenues for future research. The act of expressing one's views can lead to an internalizing effect, and the effect of feedback given by other users is limited: The only measurable influence was that a negative comment diminishes the internalization of the expressed opinion, at least when prior attitudes toward the topic are not already at the extremes of the spectrum. In a limited way, depending on the composition of online networks, this gives cause for optimism that spirals of reinforcement and polarization may be hampered by counter-speech. On the contrary, a more problematic pattern may occur when expressing opinions in a like-minded group in which feedback is predominantly positive leads to a stronger internalization of popular opinions (or when democratically valuable statements are attacked by counter-speech). It would be valuable to further investigate these processes with different topics (including hot-button issues), different sources of feedback (e.g., friends vs. unknown peers), and a consideration of long-term developments of opinion climates, repeated expressions and repeated feedback. Such a perspective on the interplay of self-effects, feedback, and reception effects would help to further advance the analysis of opinion formation in today's media landscape.

CRedit authorship contribution statement

Stephan Winter: Conceptualization, Formal analysis, Funding acquisition, Writing – original draft. **Anne L. Vos:** Conceptualization, Formal analysis, Methodology, Writing – original draft. **Paola Remmelswaal:** Conceptualization, Methodology, Writing – review & editing. **Peter Neijens:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors report there are no competing interests to declare.

Data availability

Data will be made available on request.

Acknowledgment

The present study was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG; WI 4249/2-1).

References

- Baumeister, R. F., & Tice, D. M. (1984). Role of self-presentation and choice in cognitive dissonance under forced compliance: Necessary or sufficient causes? *Journal of Personality and Social Psychology*, 46(1), 5–13. <https://doi.org/10.1037/0022-3514.46.1.5>
- Bayer, J. B., Trieu, P., & Ellison, N. B. (2020). Social media elements, ecologies, and effects. *Annual Review of Psychology*, 71(1), 471–497. <https://doi.org/10.1146/annurev-psych-010419-050944>
- Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, 6, 1–62. [https://doi.org/10.1016/s0065-2601\(08\)60024-6](https://doi.org/10.1016/s0065-2601(08)60024-6)
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brady, W. J., McLoughlin, K. L., Doan, T. N., & Crockett, M. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33). <https://doi.org/10.1126/sciadv.abe5641>
- Carr, C. T. (2021). *Computer-mediated communication: A theoretical and practical introduction to online human communication*. Rowman & Littlefield.
- Carr, C. T., & Foreman, A. C. (2016). Identity shift III: Effects of publicness of feedback and relational closeness in computer-mediated communication. *Media Psychology*, 19(2), 334–358. <https://doi.org/10.1080/15213269.2015.1049276>
- Carr, C. T., & Hayes, R. A. (2015). Social media: Defining, developing, and divining. *Atlantic Journal of Communication*, 23(1), 46–65. <https://doi.org/10.1080/15456870.2015.972282>

- Cho, J., Ahmed, S., Keum, H., Choi, Y. J., & Lee, J. H. (2018). Influencing myself: Self-reinforcement through online political expression. *Communication Research*, 45(1), 83–111. <https://doi.org/10.1177/0093650216644020>
- Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, 31, 410–436. <https://doi.org/10.1006/jesp.1995.1018>
- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook “friends”: Social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), 1143–1168. <https://doi.org/10.1111/j.1083-6101.2007.00367.x>
- Gonzales, A. L., & Hancock, J. T. (2008). Identity shift in computer-mediated environments. *Media Psychology*, 11(2), 167–185. <https://doi.org/10.1080/15213260802023433>
- Johnson, B. K., & van der Heide, B. (2015). Can sharing affect liking? Online taste performances, feedback, and subsequent media preferences. *Computers in Human Behavior*, 46, 181–190. <https://doi.org/10.1016/j.chb.2015.01.018>
- Lane, D. S., Lee, S. S., Liang, F., Kim, D. H., Shen, L., Weeks, B. E., et al. (2019). Social media expression and the political self. *Journal of Communication*, 69(1), 49–72. <https://doi.org/10.1093/joc/jqy064>
- Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin*, 9(3), 371–375. [10.1177/0146167283093007](https://doi.org/10.1177/0146167283093007)
- Lee, E.-J. (2004). Effects of visual representation on social influence in computer-mediated communication: Experimental tests of the social identity model of deindividuation effects. *Human Communication Research*, 30, 234–259. <https://doi.org/10.1111/j.1468-2958.2004.tb00732.x>
- Lee, E.-J., & Shin, S. Y. (2021). Mediated misinformation: Questions answered, more questions to ask. *American Behavioral Scientist*, 65(2), 259–276. <https://doi.org/10.1177/0002764219869403>
- Leonardi, P. M., & Treem, J. W. (2012). Knowledge management technology as a stage for strategic self-presentation: Implications for knowledge sharing in organizations. *Information and Organization*, 22(1), 37–59. <https://doi.org/10.1016/j.infoandorg.2011.10.003>
- Mitchell, A., Gottfried, J., & Matsa, K. E. Millennials and political news (Pew Research Center). Retrieved from <http://www.journalism.org/2015/06/01/millennials-political-news/>
- Nabi, R. L., Huskey, R., Nicholls, S. B., Keblusek, L., & Reed, M. (2019). When audiences become advocates: Self-induced behavior change through health message posting in social media. *Computers in Human Behavior*, 99, 260–267. <https://doi.org/10.1016/j.chb.2019.05.030>
- Nadkarni, A., & Hofmann, S. G. (2012). Why do people use Facebook? *Personality and Individual Differences*, 52(3), 243–249. <https://doi.org/10.1016/j.paid.2011.11.007>
- Nekmat, E. (2012). Message expression effects in online social communication. *Journal of Broadcasting & Electronic Media*, 56(2), 203–224. <https://doi.org/10.1080/08838151.2012.678513>
- Neubaum, G., & Krämer, N. C. (2017). Opinion climates in social media: Blending mass and interpersonal communication. *Human Communication Research*, 43(4), 464–476. <https://doi.org/10.1111/hcre.12118>
- Pingree, R. J. (2007). How messages affect their senders: A more general model of message effects and implications for deliberation. *Communication Theory*, 17(4), 439–461. <https://doi.org/10.1111/j.1468-2885.2007.00306.x>
- Rosenthal, R., & Rosnow, R. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge University Press.
- Ryff, F. A., Wirz, D. S., Kühne, R., & Wirth, W. (2014). How emotional media reports influence attitude formation and change: The interplay of attitude base, attitude certainty, and persuasion. *Media Psychology*, 17(4), 397–419. <https://doi.org/10.1080/15213269.2014.933850>
- Schlenker, B. R., Dlugolecki, D. W., & Doherty, K. (1994). The impact of self-presentations on self-appraisals and behavior: The power of public commitment. *Personality and Social Psychology Bulletin*, 20(1), 20–33. <https://doi.org/10.1177/0146167294201002>
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30(4), 526–537. <https://doi.org/10.1037/h0037039>
- Tice, D. M. (1992). Self-concept change and self-presentation: The looking glass self is also a magnifying glass. *Journal of Personality and Social Psychology*, 63(3), 435–451. <https://doi.org/10.1037/0022-3514.63.3.435>
- Toma, C. L., & Choi, M. (2015). The couple who facebooks together, stays together: Facebook self-presentation and relationship longevity among college-aged dating couples. *Cyberpsychology, Behavior, and Social Networking*, 18, 367–372. <https://doi.org/10.1089/cyber.2015.0060>
- Valkenburg, P. M. (2017). Understanding self-effects in social media. *Human Communication Research*, 43(4), 477–490. <https://doi.org/10.1111/hcre.12113>
- Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43. [10.1177/0093650996023001001](https://doi.org/10.1177/0093650996023001001)
- Walther, J. B., Kashian, N., Jang, J. W., Shin, S. Y., Dai, Y., & Koutamanis, M. (2018). The effect of message persistence and disclosure on liking in computer-mediated communication. *Media Psychology*, 21(2), 308–327. <https://doi.org/10.1080/15213269.2016.1247718>
- Walther, J. B., Liang, Y., DeAndrea, D. C., Tong, S. T., Carr, C. T., Spottswood, E. L., et al. (2011). The effect of feedback on identity shift in computer-mediated communication. *Media Psychology*, 14(1), 1–26. <https://doi.org/10.1080/15213269.2010.547832>
- Winter, S., Rimmelswaal, P., & Vos, A. (2022). When posting is believing: Adaptation and internalization of expressed opinions in social network sites. *Journal of Media Psychology*, 34(3), 177–187. <https://doi.org/10.1027/1864-1105/a000308>