

University of Groningen

## Dealing with disagreement

Roos, Carla; Koudenburg, Namkje; Postmes, Tom

*Published in:*  
New media & society

*DOI:*  
[10.1177/1461444821993042](https://doi.org/10.1177/1461444821993042)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2022

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Roos, C., Koudenburg, N., & Postmes, T. (2022). Dealing with disagreement: The depolarizing effects of everyday diplomatic skills face-to-face and online. *New media & society*, 24(9), 2153-2176.  
<https://doi.org/10.1177/1461444821993042>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



# Dealing with disagreement: The depolarizing effects of everyday diplomatic skills face-to-face and online

new media & society  
2022, Vol. 24(9) 2153–2176  
© The Author(s) 2021



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1461444821993042  
journals.sagepub.com/home/nms



Carla Anne Roos , Namkje Koudenburg  
and Tom Postmes

University of Groningen, The Netherlands

## Abstract

In online text-based discussions, people behave less diplomatically because they are more outspoken and less responsive. This can feed *impressions* of polarization. This article uses a new methodology to isolate the influence of outspokenness and responsiveness in shaping perceptions of polarization in online chat and face-to-face discussions. Text-based online and face-to-face discussions were reproduced in a face-to-face format (Study 1) and in a text-based chat format (Study 2). Uninformed observers (N = 102 and N = 103, repeated measures) evaluated these. The results showed that responsiveness was generally considered indicative of agreement and good social relationships but the interpretation of outspokenness (or lack of ambiguity) depended on the medium format. This suggests that what counts as diplomacy is not the same for each medium. Moreover, the experiences of the actors reproducing the chats in a face-to-face format highlighted the differences between media. We conclude that online conversational dynamics may play an important role in societal polarization.

## Keywords

Ambiguity, diplomacy, disagreement, online discussions, polarization, responsiveness

---

Study 2 was preregistered at [osf.io/8xfm2](https://osf.io/8xfm2)

### Corresponding author:

Carla Anne Roos, Department of Social Psychology, Faculty of Behavioral and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.

Email: [c.a.roos@rug.nl](mailto:c.a.roos@rug.nl)

Concerns about polarization within societies are mounting worldwide (Settle, 2018). Online discussions are often blamed for this because they contain rude, demeaning, and less-than-democratic comments (Coe et al., 2014; Hlavach and Freivogel, 2011), and because they might descend into conflict and polarization more easily than discussions held face-to-face (FtF) (Anderson et al., 2014). Consequently, there has been a lot of speculation about, and research into, the differences between online and FtF discussions that might help explain tendencies for conflict and polarization. Several prominent explanations focus on the characteristics of the online medium and the psychological changes this gives rise to. For example, anonymity has been argued to lead to reduced social presence (Cortese and Seo, 2012; Short et al., 1976), a state of deindividuation (Lowry et al., 2016; Sproull and Kiesler, 1986), or a reduced accountability (Santana, 2014; Suler, 2004). Alternatively, anonymity could also lead to increased group identification (Bae, 2016; Postmes et al., 1998) or more self-awareness (Nielsen, 2017). All of these have been associated with changes in the adherence to social norms (Huang and Li, 2016; Lea et al., 1992).

Recent research took a different approach by investigating what happens within the *interpersonal* dynamics of text-based online and FtF discussions (Roos et al., 2020). Instead of assuming that the online medium alters people's psychology and that this leads to behavioral changes, this research focused on the direct impact of the medium and its affordances on people's social interactions, and on the psychological effects associated with such interactions. The results showed that even in the absence of actual polarization (i.e. disagreement about the topic), online interaction partners can still have the *impression* of polarization due to an inability to enact certain diplomatic skills that they use in their FtF discussions.

However, disentangling the role of the above-mentioned factors in polarization is inherently problematic because online and FtF media differ in many ways. The first aim of this article is therefore to develop a new methodology for directly comparing discussions via different media that removes the medium as explanatory variable. By reproducing text-based online discussions in an FtF format, and reproducing FtF conversations in a text-based chat format, we aim to gain insight into the differences between online and FtF discussions. We study this qualitatively, by analyzing the challenges encountered in the process of translating discussions from one medium to the other, and quantitatively, by examining how these reproductions are perceived by outside observers. The second aim of this article is to use this methodology to empirically test the role of the diplomatic behaviors uncovered in earlier research in explaining differences in perceived polarization between media.

### *Online polarization*

Political scientists find that politics are more and more central to people's identity (Mason, 2018). As a result of intergroup social identity dynamics, political outgroups are increasingly hostile (Iyengar et al., 2012; Turner, 1991) and affective polarization occurs (Iyengar and Westwood, 2014). According to Settle (2018), social media make political and affective polarization personal: social media such

as Facebook reinforce motivated identity processes, for example, by highlighting the visibility of political conflict (intergroup threat) and by rewarding extreme utterances (reputation management). Because this occurs on social media where posts are seen by “friends”, political position and polarization make their way into social networks.

In the present article we zoom in on microdynamics: what happens *within* online and FtF discussions about politically sensitive topics. Indeed, we define polarization as an individual’s perception of the amount of disagreement about such topics *in the conversation itself*. Accordingly, we operationalize polarization as a perceived lack of shared cognition (i.e. common understanding and agreement). Importantly, these perceptions can be based on misattributions or misunderstandings of interaction partners’ motives. We believe that by taking a closer look at these microdynamics, we gain valuable insights into the origins of disagreements and their escalation. And, as we shall discuss toward the end of the article, these microdynamics may also be very relevant to macro-level political polarization.

### *Online disinhibition*

Online conversation is in many respects very different from natural FtF interaction. Most research has focused on the following distinguishing features: it is mostly a- or semi-synchronous, poor in subtle social cues, and often anonymous. These characteristics have been connected to the proneness of online interactions to polarization (Lin, 2009; Stromer-Galley and Wichowski, 2011; Suler, 2004). Most prominently, it is assumed that anonymity makes people feel more distant from others (i.e. decreased social presence, Short et al., 1976), which reduces their concerns for being positively evaluated, and motivates them to act with less inhibitions (i.e. online disinhibition, Sproull and Kiesler, 1986; Suler, 2004). In the context of contentious discussions, this disinhibition can take a toxic form, resulting in hostile and aggressive online behavior, like rude language and threats (often referred to as flaming, Lapidot-Lefler and Barak, 2012).

The evidence supporting this theory is mixed, however (Clark-Gordon et al., 2019; Lea et al., 1992; Papacharissi, 2004). While the online disinhibition thesis still receives wide support (Asker and Dinas, 2019; Lapidot-Lefler and Barak, 2012; Lowry et al., 2016), some research points to the contrary: by redirecting people’s attention from their individual identity to their social group membership, anonymity makes people *more* focused on each other and *more* motivated to adhere to social norms (Bae, 2016; Postmes et al., 1998). To explain why online discussions might more easily polarize than discussions held FtF despite people’s unabated motivation to maintain good social relations, recent research suggests that the inherent characteristics of the medium limit people’s ability to use diplomatic skills (Roos et al., 2020). This line of research suggests that the medium feeds perceptions of polarization not because the intention is different, nor because the content is more offensive, but because people cannot behave as diplomatically as they can FtF. Thus, rather than losing their *motivation* to be diplomatic online, people seem to lose their *ability* to do so.

### *Diplomatic behaviors FtF*

This research is based on a growing literature showing that FtF interaction partners often rely, consciously or not, on everyday diplomatic skills to communicate and infer the status of their relationship (e.g. Koudenburg et al., 2017; Reis and Clark, 2013). A first important skill is *responsiveness*, defined as the degree to which people provide instant feedback on each other's comments. In conversational turn-taking, responsiveness takes the form of brief signals like "yes," "hmm," or head nods during another speaker's turn or at the start of a new turn (Beňuš et al., 2011). People use responsiveness to convey that they are interested, listening, and engaged in the conversation. This signals to their partners that they are motivated to build or preserve a cooperative and pleasant relationship and to come to a common understanding and/or agreement (Davis and Perkowski, 1979; Koudenburg et al., 2017; Reis and Clark, 2013). Conversely, disrupted responsiveness, even when clearly caused by a technical failure, tends to be interpreted as a signal of disagreement and dislike (Koudenburg et al., 2013a, 2013b; Roos et al., 2020).

The second diplomatic skill of interest is *ambiguity*. Whereas communicating clearly is generally considered a good thing, it might not be in the context of contentious discussions (Brown and Levinson, 1987). In the pragmatics literature, it has been repeatedly found that people in FtF discussions prevent disagreements from escalating by ambiguating their disagreement rather than expressing it clearly (Bavelas et al., 1990; Pomerantz, 1984). To ambiguuate their expressions, people rely on disclaimers (e.g. "I do not know for sure"), hedges (e.g. "maybe," "sort of"), and vocalizations that express doubt (e.g. a drawn out "hmmm") or tentativeness (e.g. interjecting "uhmmm's") (Brennan and Clark, 1996; Reid et al., 2003). This signals that they take their partners' feelings into account and attempt to maintain a pleasant relationship with them. It also masks disagreements and allows partners to assume shared cognition (i.e. social projection, Krueger, 1998). Thus, in FtF discussions, interaction partners navigate disagreements by being responsive to each other's comments and by ambiguating their expressions, thereby maintaining a sense of solidarity and agreement.

### *Diplomatic behaviors online*

The characteristics of text-based online media could make the enactment of these diplomatic skills challenging. Roos et al. (2020) argued that the reduced synchronicity will make online conversation less responsive (i.e. comments appearing with delays and sometimes out of order), and the relative lack of subtle social cues will make online expression less ambiguous (more clear and outspoken). This could make it harder to prevent disagreements from escalating online. Indeed, comparing the way in which interaction partners handled disagreements online and FtF, the reduced responsiveness and ambiguity in text-based online discussions explained reductions in experienced shared cognition and solidarity (Roos et al., 2020). Interestingly, these studies did not find any more or less expressed disagreement or disinhibited behavior in the online chats. This suggests that, although *content-wise*, people do not express more disagreement, differences in the enactment of diplomatic skills can explain *impressions* of polarization in online discussions.

### *Varying responsiveness and ambiguity within media*

To summarize, these studies suggest that the use of everyday diplomatic skills (responsiveness and ambiguity) reduces the risk of the escalation of disagreement, and that online interactions can be more prone to polarization where these diplomatic behaviors are limited. However, this research has one intrinsic limitation that means that alternative explanations cannot be ruled out. Because the studies compare two very different communication media, other media characteristics that have also been associated with polarization may play a role. There could have been residual differences in anonymity, social presence, and so on. Although the prior research attempted to control for such confounding variables, the only way to rule out alternative explanations conclusively is to replicate the effects within the same communication medium.

Therefore, in an attempt to replicate and extend this prior research (Bonett, 2012), the present studies try to vary the conversational use of everyday diplomatic skills *within* communication media. We reproduced discussions held in text-based online chats and FtF discussions in an FtF format (Study 1) and in a text-based chat format (Study 2), and asked uninformed participants to evaluate these. By (1) keeping constant the medium in which discussions were perceived, and (2) not informing participants of the medium in which the discussions were originally held, any differences we would obtain should be attributed to the differences in text (style and content) in the “original” media. Moreover, the experience of translating a discussion from one medium (e.g. text-based chat) into another medium (FtF conversation) itself could be considered a learning process that might highlight the peculiar characteristics of, and differences between, media.

In Study 1a, we asked a group of actors to (re-)enact text-based online and FtF controversial small group discussions. The actors were interviewed about their (re-)enactment experiences. In Study 1b, we tested the hypotheses (see below) by asking individual participants to rate the video clips of these (re-)enactments on perceived responsiveness and clarity (vs ambiguity). In order to assess the social attributions following from observing these diplomatic behaviors, we also asked participants to rate the degree of shared cognition (the absence of which suggests there is perceived disagreement and lack of mutual understanding; perceived polarization) and solidarity (the absence of which suggests there is a perceived lack of good social relationships) among interaction partners. In Study 2, we took the complementary approach by asking another sample to evaluate the *transcribed* version of the same discussions.

### *Hypotheses*

In line with Roos et al. (2020), we expect that, irrespective of the communication medium format (i.e. FtF or online) in which it is perceived, participants will evaluate the original chats as less responsive (Hypothesis 1a) and more clear (Hypothesis 1b) compared to the original FtF discussions. We further hypothesize that, irrespective of the communication medium format, participants will think interaction partners in the original chats experience less shared cognition (i.e. more polarization, Hypothesis 2a) and less solidarity (Hypothesis 2b). We further hypothesize a causal relationship

between these factors, such that a reduced responsiveness and increased clarity will lead participants to conclude that interaction partners experience less shared cognition and less solidarity (Hypothesis 3a and 3b).

Alternative predictions can also be made. It seems reasonable to expect that the fit between diplomatic behaviors and medium format might play a role, in line with Expectancy Violations Theory (EVT, Burgoon and Hale, 1988; Burgoon and Walther, 1990). EVT proposes that people enter interactions with certain expectations concerning behavior, and that these expectations are shaped, among others, by communication format. People are very sensitive to violations of these expectations, and attempt to interpret and evaluate these violations as soon as they occur. Accordingly, based on prior experiences, observers might expect to see a lack of responsiveness and an abundance of clarity and outspokenness in online chats, and consider this normal and appropriate. The reverse might apply to the FtF format: here, observers would expect more responsiveness and more ambiguity. But when their expectations concerning diplomatic behaviors are violated, that is, when online conversation is responsive and ambiguous or when FtF discussion is not responsive and clear, observers might conclude that interaction partners are more polarized.

## General method

As the methods of the two studies overlap considerably, we will first describe the procedures that apply to both studies. The methodological specifics of each study will then be presented together with the results. Both studies had a within-subject design with two conditions representing different communication media sources: text-based online chat and FtF discussion. In Study 1, each participant saw one online chat and one FtF conversation in a FtF format (in randomized order: 57 first saw a FtF and 45 first saw an online discussion), and in Study 2, each participant read one online chat and one FtF conversation in an online chat format (in randomized order: 51 first saw a FtF and 52 first saw an online discussion).

### *Power and participant recruitment*<sup>1</sup>

A repeated measures within factors power calculation, with an expected effect size of  $f = .18$  (based on a similar study, Roos et al., 2020), an error probability of .05, and a desired power of .95, showed that we required 103 participants per study. Participants were recruited from a pool of Dutch first year psychology students and received partial course credit. After signing up, participants received a link to the online experiment.

### *Stimulus materials and procedures*

The stimulus materials came from a previous study in which groups of four unacquainted Dutch students discussed controversial statements: one via a text-based online chat and one FtF (Roos et al., 2020). Unknown to participants, the fourth group member was a confederate who expressed controversial views. Specifically, she spoke or wrote two scripted sentences that represented a right-wing position. To allow for a natural



conversation, the remainder of the discussion was unscripted, but the confederate was instructed to stick to her position. The text of the online discussions was saved and the FtF discussions were voice-recorded. From the resulting dataset, we selected the online chat and FtF conversations of two groups: four discussions in total.<sup>2</sup> Similar, and equivalently controversial, statements were covered in both conditions.<sup>3</sup>

In Study 1, the selected discussions were (re-)enacted by a group of actors in front of a camera. In Study 2, the same discussions were transcribed into an online chat format. The specific procedures of (re-)enacting and transcribing will be outlined below. The resulting video clips and texts were embedded in an online questionnaire so that each video clip or text was followed by the same block of questions.

Participants were unaware that the videos and texts originated from online chat and FtF interactions: participants were simply told that they would see or read two conversations about controversial topics between unacquainted students. However, in Study 1 both discussions were presented in an FtF format (i.e. in a video-recording), and in Study 2 both discussions were presented in a typical text-based chat layout.

### *Dependent measures<sup>4</sup>*

To measure perceived responsiveness, we adapted<sup>5</sup> the scale of Roos et al. (2020). Participants indicated how often (1 = *Never* to 5 = *Continuously*) they thought group members “. . . listened to each other/ . . . ignored each other (R)/ . . . cross talked (R)” ( $\alpha = .78$  in Study 1b and  $\alpha = .73$  in Study 2). In order to assess perceived clarity of expression, we developed a four-item scale that showed good reliability ( $\alpha = .80$  in Study 1b and  $\alpha = .79$  in Study 2): “In this conversation, the group members . . . expressed themselves in a clear manner/ . . . opinions were vague (R)/ . . . expressed themselves forcefully (1 = *Completely disagree/ Never* to 5 = *Completely agree/Continuously*). The fourth item was the group-level average of: “How (un)clear did the individual group members give their opinion?” (1 = *Very vague* to 7 = *Very clear*, recoded to a five-point scale).

We assessed perceived shared cognition with a three-item scale adapted from Koudenburg et al., (2013a), for example, “During this conversation, I felt that the group members understood each other” (1 = *Completely disagree* to 5 = *Completely agree*;  $\alpha = .77$  in Study 1b and  $\alpha = .75$  in Study 2).<sup>6</sup> Finally, we included a five-item perceived solidarity scale based on Koudenburg et al. (2015), for example: “During this conversation, the group members identified with each other” (1 = *Completely disagree* to 5 = *Completely agree*;  $\alpha = .85$  in Study 1b and  $\alpha = .80$  in Study 2).<sup>7,8</sup>

### *Statistical analyses*

The intraclass correlations ranged between .00 and .37 for the repeated measures of the dependent variables, suggesting clustering of scores within participants (Bliese, 2000). To account for this dependence of observations, the data were analyzed in multilevel models, where condition (Level 1) was nested in participants (Level 2). We analyzed the data with the lmer function in R (lme4, version 1.1-9, Bates et al., 2019). To test the effects of condition on perceived diplomatic behavior and perceived shared cognition



and solidarity (Hypotheses 1 and 2), we compared the fit of the multilevel models that included only the random effect of participants with the equivalent models that added condition as fixed effect predictor.<sup>9</sup> We used the emmeans package (version 1.4., Lenth et al., 2019) to estimate the condition means and confidence intervals. We tested the causal relations (Hypotheses 3) in two multilevel models with either shared cognition or solidarity as outcome variable, and both responsiveness and clarity as fixed effect predictors, nested in the random factor of participants.

## **Study 1a: performing the (re-)enactments**

Study 1 consisted of two parts. Study 1a, the qualitative and more exploratory part, covered actors' experiences through interviews before, during, and after the (re-)enactment session. Study 1b, the quantitative and hypotheses-testing part, covered observers' perceptions of these (re-)enactments as measured by the questionnaire.

### *Method study 1a*

*Procedure and stimuli.* As input material for the (re-)enactments, we retained the text-based online chats in their original state and transcribed the FtF voice-recordings in detail. We invited four young female actors from a Dutch theater group whose gender and approximate age matched the original student groups that produced the scripts (as well as the majority of the intended participant population). The (re-)enactments took place in two sessions.

In the first session, the actors were presented with the four transcripts, plus the voice-recordings of the FtF discussions, and asked how they would convert these into theatrical performances. We gave the actors a lot of freedom because we expected this would provide us with unique information. Actors are used to translating text-based scripts into FtF performances; a process in which many decisions have to be made. The peculiarities of each form of conversation might become apparent through these decisions and the difficulties encountered in their execution.

The actors wanted to learn by heart the entire chat scripts and the parts of the (relatively long) FtF conversations around the two standardized sentences of the confederate. They rotated the role of confederate among themselves. The actors decided to emphasize the differences they observed between the FtF and the online chat interactions as well as between the (in their opinion) outspoken confederate and tentative naive participants. The actors further decided to improvise the FtF filler words (e.g. "hmm," "yes") instead of keeping strictly to the scripted ones. For the online chats, the actors decided to speak out abbreviations completely, and to verbalize the typos and emoji.

The second session took place three weeks later. The actors performed the (re-)enactments seated on adjacent sides of a table in front of a camera. We edited the resulting videos to be of equal length (approximately five minutes). We selected the parts of the (re-)enactments that included the two standardized sentences of the confederate.<sup>10</sup>

*Qualitative measures.* During both sessions, we conducted semi-structured interviews in which we asked the actors about their (re-)enactment experiences and the differences they noticed between the online chat and FtF conversations. Moreover, after watching the resulting videoclips, the actors were asked about their perceptions and experiences via email. We performed an inductive thematic analysis on the actors' interview responses (see Miles and Huberman, 1994). This means that the content of their answers was analyzed to extract recurrent themes. We only extracted themes that all actors agreed on.

### *Results study 1a*

The thematic analysis revealed three themes, which will be described below with illustrative quotes.

*(Re-)Enactment challenges.* Interestingly, the actors commented not just on the difficulties of enacting the chats. They also talked about the challenges they encountered while re-enacting the FtF conversations. The quotes below illustrate the complexity for the actors.<sup>11</sup>

Q1: "It [enacting face-to-face and chat discussions] was just very different from acting a regular scene from a play." (Interview 2, Actor 1)

Q2: "A face-to-face conversation feels very uncomfortable and does not always run smoothly either, just like the chat, but only few punches are thrown." (Interview 2, Actor 1)

The actors experienced both the chat and the FtF (re-)enactments as very different from playing a regular theater script (Q1). In fact, much of the interviews revolved around the dissection of exactly what made the (re-)enactments so challenging. In both online and FtF discussions, the actors noticed discomfort evidenced by a lack of smooth flow, but the way in which they attributed this was medium dependent (Q2): disagreement leads to unpleasant conversations, no matter in which medium it is encountered, but only the interaction partners in the chat discussions confronted this aggressively. In FtF discussions, people refrained from seeking confrontations ("only few punches are thrown"), but clearly showed their discomfort while doing so. The latter might fulfill a social function by communicating tact and good intentions (cf. politeness, Brown and Levinson, 1987). It thus appears that the factors making (re-)enactments challenging differ between media. The two other themes uncover these different factors.

*Empty phrases.* The main challenge encountered in the FtF re-enactments were the many "empty" or "content-less" phrases and utterances, as the actors called them, such as "uhm" and "or so," that were absent in the online chats. The following quotes illustrate how the actors saw this difference between media and its consequences.

Q3: “All the statements in the chat conversations are clear, the sentences are clear, while the face-to-face conversations are often more nuanced through the use of words like ‘say’, ‘or something’, ‘I think’.” (Interview 3, Actor 1)

Q4: “The word choice and the doubtful replies in the face-to-face conversations show that people are probing what they can and cannot say in this group. Judging by their replies, the chatters seem to take that less into account.” (Interview 3, Actor 1)

Q5: “I think that the people in the face-to-face conversation might have felt quite tense. I do not think they knew each other. There is always a social aspect involved: you want to be liked or maintain a good atmosphere. That results in very tentative expressions of opinions. The chat is safer, because you do not see each other. It is easier to simply say what you think. I think that those people felt more at ease.” (Interview 3, Actor 2)

Q6: “Speaking in the chat is a lot less nuanced and takes on the form of rounded ‘statements.’ These do not connect as well as expressions in a face-to-face conversation. However, it also struck me that those without an extreme opinion in the face-to-face discussion had some difficulty with formulating their thoughts. A lot of words were spoken, but no clear point was made. That made the discussion woolly and more difficult to enact.” (Interview 3, Actor 1)

To the actors, the chat conversations seemed much more clear, efficient, and to-the-point than the FtF conversations, in which people expressed themselves rather ambiguously and used many more words than necessary. Quotes 4 to 6 show that these differences shaped the actors’ impressions of interaction partners’ intentions. The actors thought the interaction partners in the FtF discussions were tentatively exploring what was deemed acceptable conduct in order “to be liked or maintain a good atmosphere” (Q4 and 5). Thus, FtF people seemed to be more concerned with each other’s opinion than with their own. On the contrary, actors thought that the interaction partners in the online chats did not care much about each other’s opinion, and, consequently, were less afraid for social rejection and less reluctant to share their opinion. Another reason the actors mentioned for the apparent ease with which online interaction partners expressed themselves was being convinced and holding a clear opinion. Conversely, the empty phrases in the FtF conversations were interpreted as a process of formulating thoughts and therefore indicative of a lack of an “extreme opinion” (Q6). Interestingly, the undecided “wooliness” of the FtF discussions made these harder to re-enact. The actors found it easier to remember their lines when their characters make a clear point. Indeed, the actors tried to define their character based on his or her statements, which were rather ambiguous in the FtF discussions.

*Action–reaction.* In enacting the online chats, the actors especially noticed the difficulty they had in suppressing their natural tendency to respond to each other’s comments, a challenge that was absent in the FtF re-enactments. The quotes below illustrate how the actors experienced this difference between media and its consequences.

Q7: “In acting, and in normal conversations, you want to be able to react to each other. Action-reaction. That was possible in the face-to-face conversations. In the chats that was difficult.” (Interview 3, Actor 2)

Q8: “If you concentrate on what is actually being said, you notice that the chat is like a weird robot interaction, as if people are not really listening to each other. The face-to-face conversation comes across as being a lot more natural! That kind of surprises me, it was difficult for us to incorporate so many ‘loose’ words but it worked and that is nice to see.” (Interview 3, Actor 1)

Q9: “People tried to maintain a good atmosphere in the face-to-face conversations. They also tried to listen to each other and understand each other, what happened less in the chats. There [in the chats], people sometimes completely talked past each other because everyone was mainly concerned with sharing his/her own opinion. The atmosphere was virtually unimportant in the chats.” (Interview 3, Actor 2)

The actors thought that the increased to-the-point-ness in the online discussions could explain why interaction partners reacted to each other less (Q6). Actors repeatedly characterized the online chat as a “weird robot interaction” (Q8). Just like robots, interaction partners appeared unaware of each other’s presence: rather than listening to each other, they seemed to be more concerned with venting their own opinion. This resulted in a disjointed conversation where interaction partners appeared to be unconcerned with, and even dislike, each other (Q9). Conversely, the actors thought the many empty phrases in the FtF discussions promoted natural action–reaction sequences (Q8). Actors saw this responsiveness as a sign of interaction partners being concerned with each other and motivated to maintain good relations (Q9). Note the overlap with the impressions resulting from the (lack of) empty phrases.

*Connecting empathy phrases with action–reaction.* To conclude, both the online chats and the FtF discussions were challenging to (re-)enact, but for different reasons: the FtF discussions contained many “empty” phrases and the online chats were lacking smooth action–reaction sequences. Interestingly, these characteristics appear tightly interconnected: the clear to-the-point chat sentences do not connect well, and the empty FtF phrases promote smooth action–reaction sequences. Apparently, even though both felt awkward to the actors, the latter combination resulted in a more natural conversation. Furthermore, both characteristics contributed to the actors’ perceptions of the way interaction partners related to each other. The online combination of clear and disjointed sentences communicates that interaction partners care more about their own opinion than about each other more about their own opinion than about each others’s feelings. Conversely, the FtF combination of empty phrases and coherent action–reaction signals that interaction partners prioritize their relationship.

## Study 1b: observing the (re-)enactments

### Method study 1b

**Participants and design.** A sample of 102 participants ( $M_{\text{age}} = 19.8$ ,  $SD_{\text{age}} = 2.0$ ; 78.4% female) viewed the video clips of the (re-)enactments and rated these on responsiveness, clarity, shared cognition, and solidarity. The political orientation of the sample was skewed to the right with 54.9% describing themselves as falling between extreme and moderate left-wing, 18.6% placing themselves exactly in the middle, and 26.5% allocating themselves between moderate and extreme right on the political scale. Most participants were non-religious (82.4%; the religious participants were either Christian, Protestant, or Catholic).<sup>12</sup>

### Results Study 1b<sup>13</sup>

We found medium to large condition effects on the four dependent variables supporting Hypotheses 1 and 2 (Table 1). Specifically, participants felt interaction partners were less responsive ( $d = 0.58$ ) and expressed themselves clearer ( $d = -1.33$ ), and experienced less shared cognition ( $d = 0.81$ ) and solidarity ( $d = 0.90$ ) in the enacted online chats than in the re-enacted FtF discussions.

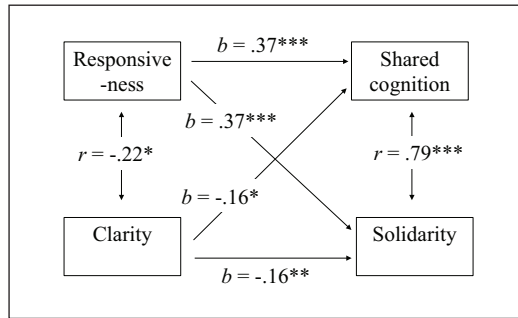
The results of the regression analyses are shown in Figure 1. In line with Hypothesis 3, responsiveness positively predicted perceived shared cognition and solidarity (both  $b = .37$ ,  $p < .001$ ), and clarity negatively predicted perceived shared cognition and solidarity ( $b = -.16$ ,  $p = .015$  and  $b = -.16$ ,  $p = .008$ , respectively). In other words, the less responsive and clearer observers thought interaction partners expressed themselves, the less they thought interaction partners were in agreement and feeling connected. As the intercorrelation between responsiveness and clarity was weak ( $r = -.22$ ,  $p = .025$ ), we can assume that their effects are largely unique.

**Table 1.** For each dependent variable measured in Study 1b, the test results of the difference between conditions (chi-square test and effect size), and the means with 95% confidence intervals per condition.

	$\chi^2(1)$	Re-enacted FtF M [95% CI]	Enacted chats M [95% CI]	$d^a$
Responsiveness	28.76***	4.04 [3.87, 4.21]	3.38 [3.21, 3.55]	0.58
Clarity	75.37***	2.83 [2.69, 2.97]	3.77 [3.63, 3.91]	-1.33
Shared cognition	31.30***	2.89 [2.74, 3.05]	2.24 [2.09, 2.40]	0.81
Solidarity	38.23***	3.23 [3.08, 3.37]	2.55 [2.41, 2.70]	0.90

Note. \*\*\* $p < .001$ .

<sup>a</sup>Cohen's  $d$  was calculated by subtracting the chat estimates from the FtF estimates and dividing this by the total SD of the full model (Cohen, 1988).



**Figure 1.** Repeated measures regressions where perceived responsiveness and clarity predict perceived shared cognition and solidarity in Study 1b (unstandardized betas are reported). The repeated measures correlations between responsiveness and clarity, and between shared cognition and solidarity are also shown.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

## Discussion Study 1

In this study, discussions that were originally held via text-based chat or FtF were all (re-)enacted in an FtF setting. Observers (and actors) noticed the differences between conditions as predicted. The original online discussions were considered less responsive and clearer than the original FtF discussions. Furthermore, to observers (and actors), online interaction partners seemed to be more concerned with clearly venting their own opinion than with listening to each other and therefore seemed to disagree more and experience less solidarity. The results are in line with the hypothesis that regardless of the medium, responsiveness and ambiguity in controversial discussion are seen as indicators of agreement and good social relations. However, these findings might instead be explained by a naturally better fit between these diplomatic behaviors and the FtF medium. To test whether the observed effects are truly independent of the medium, we ran a second study in which we assessed how observers perceive these diplomatic behaviors in a text-based online chat.

## Study 2: observing the transcripts

### Method Study 2

**Participants and design.** In Study 2, we presented another sample of participants with the textual versions of the chat and FtF discussions used in Study 1. We transcribed the parts of the discussions that actors (re-)enacted, which made the transcripts of both conditions approximately equal in length. We edited the discussions as little as possible. All edits were aimed at improving the readability and/or comparability of the transcripts. When in doubt, we used a conservative principle, that is, differences between media were minimized rather than maximized. The chat transcripts were kept almost unedited. Only the characteristics that could not be present FtF, and might therefore confound the results, were edited: we removed spelling errors (language errors were maintained), replaced the

one emoji by “haha,” excluded parentheses, and wrote out abbreviations. The FtF conversations were transcribed by removing unfinished sentences (as these might unjustly raise the impression that speakers were interrupted), writing out overlapping speech as if comments appeared after each other (but not when this would unjustly raise the impression of ignoring), adding commas and some periods to indicate speaking pauses, adding ellipses (. . .) when a speaking turn was ended prematurely, including citation marks (“”) where people referred to a specific term or cited someone else, and replaced laughing by “haha” (not for chuckling as this might unjustly communicate sarcasm). The FtF transcripts were structured like the original online chats where four differently colored blocks with a letter (*M*, *P*, *R*, or *H*) represented the four group members (see Figures 3 and 4 for an example of a FtF and a chat transcript, respectively).

A sample of 103 participants ( $M_{\text{age}} = 19.6$ ,  $SD_{\text{age}} = 1.9$ ; 82.5% female) read the transcripts and rated these on responsiveness, clarity, shared cognition, and solidarity. The political orientation of the sample was skewed to the right: 59.2% extreme to moderate left-wing, 24.3% middle, and 16.5% moderate to extreme right-wing. Most participants were non-religious (86.4%); religious participants were either Christian, Protestant, or Catholic).

## Results Study 2

Condition effects were partially different from those observed in Study 1b (see Table 2). In line with Hypothesis 1b (and Study 1b), participants considered the online chat transcripts clearer than the FtF transcripts ( $d = -1.40$ ). However, the data did not support Hypothesis 1a: condition had no statistically significant effect on perceived responsiveness ( $d = -0.01$ ). Finally, although the difference was not significant for shared cognition ( $d = -0.16$ ) and only borderline significant for solidarity ( $d = -0.22$ ), the trend contradicted Hypothesis 2 (and Study 1b): participants observed *more* rather than less shared cognition and solidarity among interaction partners in the transcribed online chats than in the transcribed FtF discussions.

Whereas we did not observe significant differences on responsiveness and perceived shared cognition between conditions, it is still interesting to test whether (within medium) variations in perceived responsiveness and clarity can explain variations in perceived shared cognition and solidarity. In line with Hypothesis 3a (and Study 1b), responsiveness positively predicted perceived shared cognition and solidarity ( $b = .21$ ,  $p = .005$  and  $b = .26$ ,  $p < .001$ , respectively, see Figure 2). Contradicting Hypothesis 3b (and Study 1b), clarity did not significantly predict perceived shared cognition, and *positively* predicted perceived solidarity ( $b = .03$ ,  $p = .604$  and  $b = .13$ ,  $p = .010$ , respectively). A statistically insignificant correlation ( $r = .12$ ,  $p = .211$ ) between responsiveness and clarity suggests that their effects are independent.

## Discussion Study 2

The results of Study 2 partially support the hypotheses *and* an alternative Medium-Fit argument. In line with the hypotheses, the discussions that were originally held FtF made an ambiguous impression on observers even when they were transcribed. However, condition did not have a significant effect on perceived responsiveness, implying that the behaviors that make FtF discussions more responsive (e.g. interjections such as “yes”



and numerous other vocalizations) are not experienced in the same way when they are written out. This suggests that expectations for responsiveness in textual exchanges are different. There was also no effect on shared cognition, which is inferred in part from responsiveness. In fact, there was a trend toward *more* shared cognition and (even significantly) *more* solidarity in the original chats, which contradicts the hypotheses but suggests a Medium-Fit effect.

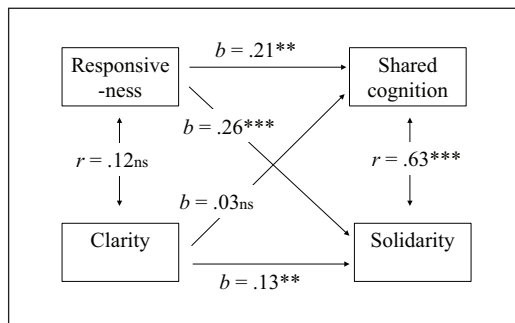
When examining the results of the regressions, we see a similar pattern of support for some hypotheses and not for others. With respect to responsiveness, the more there was of it, the more agreement and solidarity observers inferred. But clarity was *positively* associated with solidarity (instead of negatively, as predicted). This may also be explained by a Medium-Fit argument: in textual interactions, observers expect outspokenness and therefore interpret the FtF ambiguity negatively (e.g. as evasiveness), concluding that social relationships are compromised.

**Table 2.** For each dependent variable measured in Study 2, the test results of the difference between conditions (chi-square test and effect size), and the means with 95% confidence intervals per condition.

	$\chi^2(1)$	Transcribed FtF M [95% CI]	Transcribed chats M [95% CI]	$d^a$
Responsiveness	0.01 <sup>ns</sup>	3.75 [3.60, 3.90]	3.76 [3.61, 3.91]	-0.01
Clarity	112.40 <sup>***</sup>	2.50 [2.37, 2.63]	3.59 [3.47, 3.72]	-1.40
Shared cognition	1.86 <sup>ns</sup>	2.79 [2.63, 2.95]	2.94 [2.78, 3.10]	-0.16
Solidarity	4.59 <sup>*</sup>	2.94 [2.81, 3.07]	3.13 [3.00, 3.26]	-0.22

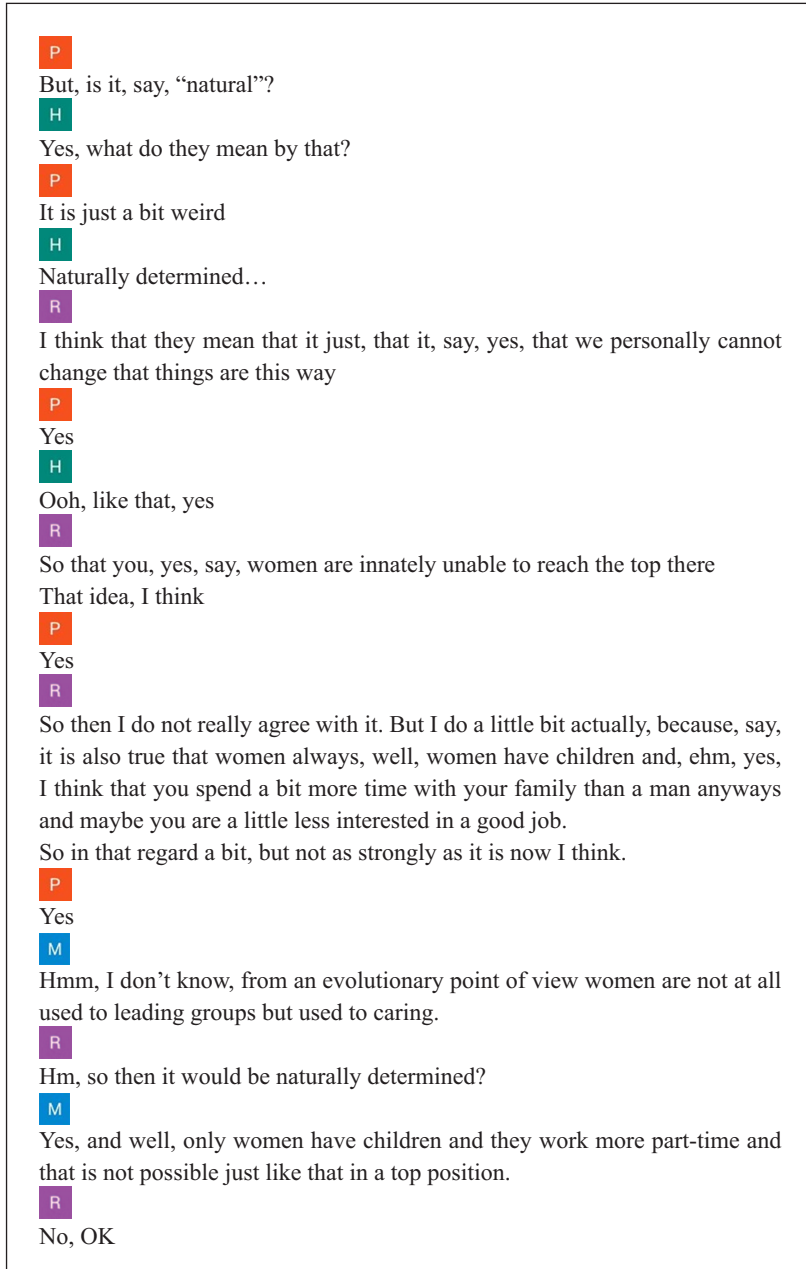
<sup>ns</sup> $p > .05$ , <sup>\*</sup> $p < .05$ , <sup>\*\*\*</sup> $p < .001$ .

<sup>a</sup>Cohen's  $d$  was calculated by subtracting the chat estimates from the FtF estimates and dividing this by the total standard deviation of the full model (Cohen, 1988).

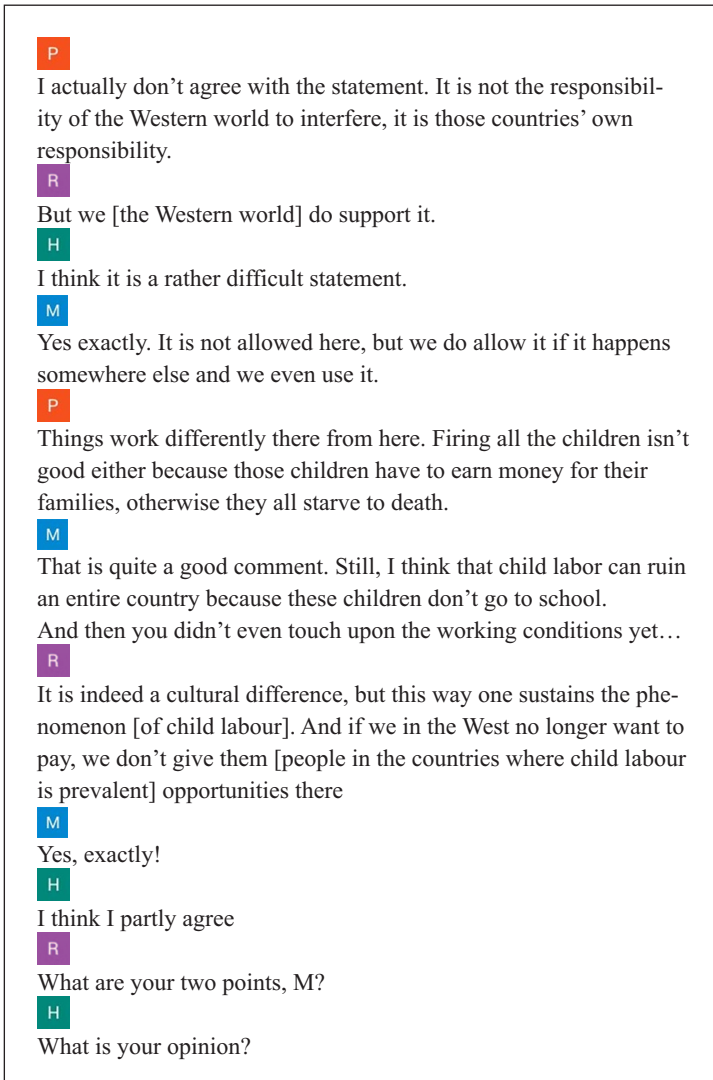


**Figure 2.** Repeated-measures regressions where perceived responsiveness and clarity predict perceived shared cognition and solidarity in Study 2 (unstandardized betas are reported). The repeated measures correlations between responsiveness and clarity, and between shared cognition and solidarity are also shown.

<sup>ns</sup> $p > .05$ , <sup>\*\*</sup> $p < .01$ , <sup>\*\*\*</sup> $p < .001$ .



**Figure 3.** First section of one of the two face-to-face conversations presented to participants in Study 2. The discussed statement is “It is natural that there are few women in the top of business and government.” The two standardized comments by the confederate are included (person M). Translated to English by the first author.



**Figure 4.** First section of one of the two chat conversations presented to participants in Study 2. The discussed statement is “Dutch businesses that support child labor should be punished heavily.” The two standardized comments by the confederate are included (person P). Translated into English by the first author. The text between brackets was not included in the stimulus materials but added to this figure for ease of reading.

## General discussion

It is often said that online discussions are more prone to polarization than FtF discussions. Recent research suggests that the subtle skills of everyday diplomacy play a central role (Roos et al., 2020). Even when there is no real polarization in terms of

disagreement, online interaction partners can still perceive polarization because contention is dealt with less responsively and disagreement is conveyed more clearly. However, as these conclusions were drawn from between media comparisons, other media characteristics, such as anonymity, may have confounded these effects. Therefore, the current article developed a new methodology to test the generalizability of these findings from between media to within media. By keeping constant the medium in which discussions were perceived and assessing the perceptions of observers that were not aware of the medium in which the discussions were originally held, we isolated the consequences of diplomatic behaviors.

### *Translating chat to FtF*

First of all, the process of translating discussions from one communication medium to the other proved very informative about the differences between media. In a thematic analysis of the actors' reflections on the (re-)enactments, we inductively identified three (recurring) themes: (1) the challenging experience of (re-)enacting both chat and FtF discussions, (2) the many empty phrases in FtF discussions, and (3) the lack of action-reaction sequences in online chats.

The (re-)enactments of both the chat and the FtF discussions were challenging, but both in their unique ways. The re-enactment of FtF discussions felt unnatural to actors mostly because of the amount of content-less speech. This took the form of short interjected expressions like "hmm" and "or so," but also of long speaking turns in which it did not become clear what interaction partners actually meant or wanted to say. While reading and enacting, actors felt these "empty phrases" made the interaction feel uncomfortable and weird, but when looking at the resulting videoclips, actors saw that these fulfilled an important social function: interaction partners seemed to be holding back to first explore what was acceptable to say before sharing their thoughts. In the online chats, actors noticed that interaction partners were much more to-the-point, which disrupted the smooth transition between speakers and gave these conversations an unnatural quality too. Interaction partners seemed to be dedicated to clearly communicating their own opinion, but to ignore each other's comments in the process. Interestingly, these observations support the hypotheses that a reduced responsiveness and increased clarity lead to perceptions of disrupted social relationships.

### *Observers' perceptions*

The hypotheses were tested by having participants rate original online chat and FtF discussions in an FtF (Study 1b) and online chat format (Study 2). Together, the findings partially support the hypotheses (based on Roos et al., 2020) and partially support an alternative Medium-Fit argument (based on Expectancy Violation Theory, Burgoon and Hale, 1988; Burgoon and Walther, 1990).

### *FtF and text-based responsiveness*

In line with Hypothesis 3a, in both FtF and text-based chat formats, observers considered responsiveness indicative of agreement and solidarity. However, what is considered

responsive behavior appears to be medium-specific. Echoing the disrupted action–reaction sequences noticed by the actors, observers thought the original online chats were less responsive, and therefore more divided, when seeing these enacted FtF. But when presented with these same online discussions in a chat format, observers did not consider these any less, nor more, responsive than the original FtF discussions. This may point to a Medium-Fit effect: the behaviors that make FtF discussions more responsive might be experienced as disrupting the smooth coordination of online chats, thereby reducing responsiveness. Considering the importance of responsiveness in fostering a sense of common understanding and agreement (e.g. Davis and Perkowski, 1979; Reis and Clark, 2013), this absence of a condition effect on responsiveness could also explain the absence of an effect on perceived shared cognition in the transcripts.

### *FtF and text-based clarity*

In line with Hypothesis 1b, independently of the medium format in which it was presented, observers thought interaction partners in the original online discussions expressed themselves relatively clearly. However, in line with a Medium-Fit argument, the consequences of clarity depended on the medium format. In the enactments, observers (and actors) felt clarity signaled disrupted social relations. But in the transcripts, clarity was associated with slightly *increased* solidarity and unrelated to shared cognition. This suggests that in online chats, it is more acceptable to provide an unvarnished and outspoken opinion than it is FtF. Conversely, the FtF ambiguity—the many hesitations, qualifiers, and hedges (such as “ehm,” “maybe,” and “sort of”)—when written out in text might be experienced weird and confusing. Consequently, observers may think something is wrong (see also Koudenburg et al., 2013a, 2017): Do interaction partners disagree? Is their relationship under threat? Thus, whereas outspokenness might be experienced negatively when comparing online with FtF media, as Roos et al. (2020) observed, it might be perceived positively when considered *within* online media, as the current findings suggest.

This can be connected to the observation that the behaviors that make FtF discussions responsive might not work online. Indeed, ambiguity in the enactments was considered conducive to responsiveness by actors (empty phrases promoting smooth action–reaction sequences) and observers alike (negative correlation between clarity and responsiveness). However, this relationship was absent in the transcripts. This suggests that ambiguity serves a responsiveness function FtF, but that ambiguating online might be experienced as avoiding the issue and therefore not responsive at all.

In sum, the findings of Roos et al. (2020) were replicated in the context of FtF discussions but not in the text-based chat format. This suggests that what counts as diplomacy is not the same for each medium. Responsiveness seems to be a diplomatic skill that helps preserve good social relations in both online and FtF discussions. But, in line with a Medium-Fit argument, observers appear to (unwittingly) prefer a match between the degree of outspokenness and the communication medium. That is, whereas ambiguity can be used to good effect to prevent escalation in FtF discussions, ambiguity in online chats may be interpreted as evasiveness and thus backfire.

### *Limitations and future directions*

In the process of translating discussions held in one medium into another medium, many, often rather subjective, decisions had to be made. For example, the actors added some nonverbal cues (such as eye contact) to the purely textual chat discussions, and we did not include all interruptions when transcribing the FtF discussions. These adaptations were made to prevent the conversations from becoming completely incomprehensible or clearly unnatural, which would introduce extra noise. These decisions underscore just how different these media are and how difficult it is to directly compare them. However, we believe that there is still value in attempting to do so, precisely because we want to learn more about the differences between media. Moreover, the adaptations we made, in effect making chat and FtF more similar, would undermine the true differences between conditions rather than artificially increasing them. In fact, our conservative approach could be an alternative explanation for the absence of a main effect on responsiveness in Study 2. Further research could explore the consequences of making different choices in the translation processes.

Although the used method allows us to keep constant many characteristics that differentiate media, the conditions arguably still differed on other characteristics besides responsiveness and outspokenness, most importantly expressed disagreement. We can assume these differences are negligible, however, because the discussions used as stimulus materials came from a previous study in which no medium effect on expressed disagreement was found.

Relatedly, as the stimulus materials were controversial discussions held in small groups of strangers in a lab environment, the generalizability of the findings might be limited to this specific type of discussions. It could be that the observed processes work out differently in large-scale discussions, in non-contentious conversations, or in discussions among non-strangers. However, by pointing away from anonymity as a key driver of online polarization effects, our findings could also help explain why polarization may occur in contexts such as social media platforms, in which people are not anonymous to each other, but are limited in the diplomatic skills they can enact. We therefore encourage future research to assess replicability in different online contexts.

The generalizability of the results is further limited by our sample: young, Western, and highly educated participants. Studying young adults, who are relatively accustomed to participating in online conversations, can be considered a conservative test of our hypotheses: more senior people's expectations about diplomatic behaviors in online discussions might more closely match their norms for FtF discussions. As a result, more senior people might consider the online lack of responsiveness and ambiguity more indicative of polarization. Furthermore, different cultures are expected to have differing norms concerning diplomatic behavior (Culpeper et al., 2010). This might also be the case for people from different educational levels and socioeconomic backgrounds. These are all interesting directions for future research.

### **Conclusion**

The present research shows that polarization can occur merely by the way in which text-based online media lead people to express themselves and by how audiences

subsequently interpret their messages. Importantly, none of these effects were caused by selective exposure due to the algorithms, filters, or bubbles, which according to other scholars are key drivers of polarization (e.g. Settle, 2018). Rather, polarization can be perceived simply due to the misattribution of unintentional behavior. Arguably, the piling up of these misunderstandings on a micro-level might lead to the formation of camps that feel ever more divided, resulting in a polarized society on the macro-level. That is, the seeds for societal polarization might be sown on this micro-level, unknowingly, unintentionally.

Accordingly, we argue that to understand polarization we should look beyond the dynamics of “us” and “them” and people’s perceptions of society at large, and move beyond the use of big data and large-scale surveys. We believe that to understand how polarization occurs in the midst of communities both online and offline, the dynamics that lead to escalation and de-escalation need to be studied more closely. By presenting a new method for comparing these dynamics within different media, we hope to inspire research that aims to uncover, and ultimately promote, dynamics that contribute to an undivided and harmonious society.

### Acknowledgements

The authors thank the four actors of theater group Vals Alarm for performing the (re-)enactments and sharing their experiences with us: Margot Vlot, Vera van Burgsteden, Willemijn Bouma, and Yteke Hoekstra. The authors additionally thank Justin Richardson for helping with the translation of the actors’ quotes to English.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Carla Anne Roos  <https://orcid.org/0000-0002-0549-1342>

### Notes

1. Both studies were approved by the Ethical Committee Psychology of the University of Groningen. Pre-participation informed consent was obtained from all participants.
2. Both selected groups showed the condition effect on responsiveness and clarity as reported earlier. The selection was further limited by the requirements that (1) the original group members were unacquainted (as this might strongly affect their manner of conversing) and (2) the order of conditions was balanced between groups.
3. The statements discussed FtF were: “It is natural that there are few women in the top of business and government” and “People whose asylum claims have been dismissed are entitled to food, drink, and shelter.” The statements discussed in the chats were: “Dutch businesses that support child labor should be punished heavily” and “No more mosques should be built in the Netherlands.”
4. All items were presented to participants in Dutch.
5. The items were adapted to the context of observers by rephrasing “we” to “the interaction partners.”



6. Koudenburg et al. (2013a) used two more items (“I had the feeling my opinions were validated” and “I had the feeling my opinions were shared”) that did not fit the current (observer) context.
7. Koudenburg et al. (2015) broadly defined solidarity and measured it with 22 items, of which we selected five items that we deemed relevant in the context of the current study.
8. The questionnaire contained a couple of additional items for exploratory purposes. The entire questionnaire is available from the corresponding author on request.
9. We specified fixed effects predictors because we expect the effects to be similar across participants.
10. The resulting videoclips are available from the corresponding author on request.
11. The quotes were literally translated from Dutch to English. The text between brackets was not mentioned by actors but added here to promote understanding of the quotes.
12. One Muslim participant was discarded because one of the discussions concerned mosques.
13. The datasets of Studies 1 and 2 are available from the corresponding author on request.

## References

- Anderson AA, Brossard D, Scheufele DA, et al. (2014) The “nasty effect”: online incivility and risk perceptions of emerging technologies. *Journal of Computer-mediated Communication* 19(3): 373–387.
- Asker D and Dinas E (2019) Thinking fast and furious: emotional intensity and opinion polarization in online media. *Public Opinion Quarterly* 83(3): 487–509.
- Bae M (2016) The effects of anonymity on computer-mediated communication: the case of independent versus interdependent self-construal influence. *Computers in Human Behavior* 55: 300–309.
- Bates D, Maechler M, Bolker B, et al. (2019) *Linear Mixed-effects Models Using “Eigen” and S4 (Version 1.1-21)*. Available at: <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- Bavelas JB, Black A, Chovil N, et al. (1990) *Equivocal Communication*. New York: SAGE.
- Beňuš s, Gravano A and Hirschberg J (2011) Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics* 43(12): 3001–3027.
- Bliese PD (2000) Within-group agreement, non-independence, and reliability: implications for data aggregation and analysis. In: Klein KJ and Kozlowski SWJ (eds.) *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco, CA: JosseyBass, pp. 349–381.
- Bonett D (2012) Replication-extension studies. *Current Directions in Psychological Science* 21(6): 409–412.
- Brennan SE and Clark HH (1996) Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning Memory and Cognition* 22(6): 1482–1493.
- Brown P and Levinson SC (1987) *Politeness: Some Universals in Language Usage (Studies in Interactional Sociolinguistics, 4)*. Cambridge: Cambridge University Press.
- Burgoon JK and Hale JL (1988) Nonverbal expectancy violations: model elaboration and application to immediacy behaviors. *Communication Monographs* 55(1): 58–79.
- Burgoon JK and Walther JB (1990) Nonverbal expectancies and the evaluative consequences of violations. *Human Communication Research* 17(2): 232–265.
- Clark-Gordon CV, Bowman ND, Goodboy AK, et al. (2019) Anonymity and online self-disclosure: a meta-analysis. *Communication Reports* 32(2): 98–111.
- Coe K, Kensi K and Rains SA (2014) Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication* 64(4): 658–679.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. Abingdon: Routledge.

- Cortese J and Seo M (2012) The role of social presence in opinion expression during FtF and CMC discussions. *Communication Research Reports* 29(1): 44–53.
- Culpeper J, Marti L, Mei M, et al. (2010) Cross-cultural variation in the perception of impoliteness: a study of impoliteness events reported by students in England, China, Finland, Germany and Turkey. *Intercultural Pragmatics* 7; 4: 027.
- Davis D and Perkowski WT (1979) Consequences of responsiveness in dyadic interaction: effects of probability of response and proportion of content-related responses on interpersonal attraction. *Journal of Personality and Social Psychology* 37(4): 534–550.
- Hlavach L and Freivogel WH (2011) Ethical implications of anonymous comments posted to online news stories. *Journal of Mass Media Ethics* 26(1): 21–37.
- Huang G and Li K (2016) The effect of anonymity on conformity to group norms in online contexts: a meta-analysis. *International Journal of Communication* 2016; 10.
- Iyengar S and Westwood S (2014) Fear and loathing across party lines: new evidence on group polarization. *American Journal of Political Science* 59; 690–707.
- Iyengar S, Sood G and Lelkes Y (2012) Affect not ideology: a social identity perspective on polarization. *Public Opinion Quarterly* 76(3): 405–431.
- Koudenburg N, Postmes T and Gordijn EH (2013a) Conversational flow promotes solidarity. *PLOS ONE* 8(11): 1–6.
- Koudenburg N, Postmes T and Gordijn EH (2013b) Resounding silences: subtle norm regulation in everyday interactions. *Social Psychology Quarterly* 76(3): 224–241.
- Koudenburg N, Postmes T and Gordijn EH (2017) Beyond content of conversation. *Personality and Social Psychology Review* 21(1): 50–71.
- Koudenburg N, Postmes T, Gordijn EH, et al. (2015) Uniform and complementary social interaction: distinct pathways to solidarity. *Plos ONE* 10(6): 1–29.
- Krueger J (1998) On the Perception of Social Consensus. *Advances in Experimental Social Psychology* 30: 163–240.
- Lapidot-Lefler N and Barak A (2012) Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28(2): 434–443.
- Lea M, O'Shea T, Fung P, et al. (1992) “Flaming” in computer-mediated communication: observations, explanations, implications. In: Lea M (ed) *Contexts of Computer-mediated Communication*. Birmingham: Harvester, pp. 89–112.
- Lenth R, Singmann H, Love J, et al. (2019) *Estimated Marginal Means, Aka Least-squares Means (Version 1.4.1)*. Available at: <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>
- Lin CA (2009) The effects of the Internet. In: Bryant J and Oliver MB (eds) *Media Effects: Advances in Theory and Research*, 3rd ed. Abingdon: Routledge, pp. 567–591.
- Lowry P, Zhang J, Wang C, et al. (2016) Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning (SSSL) model. *Information Systems Research* 27(4): 962–986.
- Mason L (2018) *Uncivil Agreement: How Politics Became Our Identity*. Chicago, IL: University of Chicago Press.
- Miles MB and Huberman AM (1994) *Qualitative Data Analysis: An Expanded Sourcebook*. New York: Sage.
- Nielsen MISW (2017) Computer-mediated communication and self-awareness—A selective review. *Computers in Human Behavior* 76: 554–560.
- Papacharissi Z (2004) Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media and Society* 6(2): 259–283.
- Pomerantz A (1984) Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In: Atkinson M and Heritage J (eds) *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, pp. 57–101.

- Postmes T, Spears R and Lea M (1998) Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research* 25(6): 689–715.
- Reid SA, Keerie N and Palomares NA (2003) Language Gender Salience, and Social Influence. *Journal of Language and Social Psychology* 22(2): 210–233.
- Reis HT and Clark MS (2013) Responsiveness. In: Simpson JA and Campbell L (eds) *The Oxford Handbook of Close Relationships*. Oxford: Oxford University Press, pp. 400–423.
- Roos CA, Koudenburg N and Postmes T (2020) Online social regulation: when everyday diplomatic skills for harmonious disagreement break down. *Journal of Computer-Mediated Communication* 25(6): 382–401.
- Roos CA, Postmes T and Koudenburg N (2020) The microdynamics of social regulation: comparing the navigation of disagreements in text-based online and face-to-face discussions. *Group Processes & Intergroup Relations* 23(6): 902–917.
- Santana A (2014) Virtuous or vitriolic: the effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice* 8(1): 18–33.
- Settle JE (2018) *Frenemies: How Social Media Polarizes America*. Cambridge: Cambridge University Press.
- Short J, Williams E and Christie B (1976) *The Social Psychology of Telecommunications*. Hoboken, NJ: John Wiley & Sons.
- Sproull L and Kiesler S (1986) Reducing social context cues: electronic mail in organizational communication. *Management Science* 32(11): 1492–1512.
- Stromer-Galley J and Wichowski A (2011) Political discussion online. In: Consalvo M and Ess C (eds) *The Handbook of Internet Studies*. New York: Wiley-Blackwell, pp. 168–187.
- Suler J (2004) The online disinhibition effect. *Cyberpsychology & Behavior* 7(3): 321–326.
- Turner JC (1991) *Social influence*. Pacific Grove, CA: Thomson Brooks/cole Publishing Co.

### Author biographies

Carla Anne Roos is a PhD student at the Social Psychology department of the University of Groningen. She studies the social regulation of disagreements in text-based online and face-to-face discussions.

Namkje Koudenburg is an associate professor at the Social Psychology department of the University of Groningen. Her research focuses on the social consequences of conversational dynamics.

Tom Postmes is a full professor at the Social Psychology department of the University of Groningen. He studies group processes within and between groups, social identity, and social norms.