university of groningen

University Medical Center Groningen

University of Groningen

Everyday Diplomacy

Roos, Carla

DOI:
10.33612/diss.230455324

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. (2022). *Everyday Diplomacy: dealing with controversy online and face-to-face*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. https://doi.org/10.33612/diss.230455324

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.

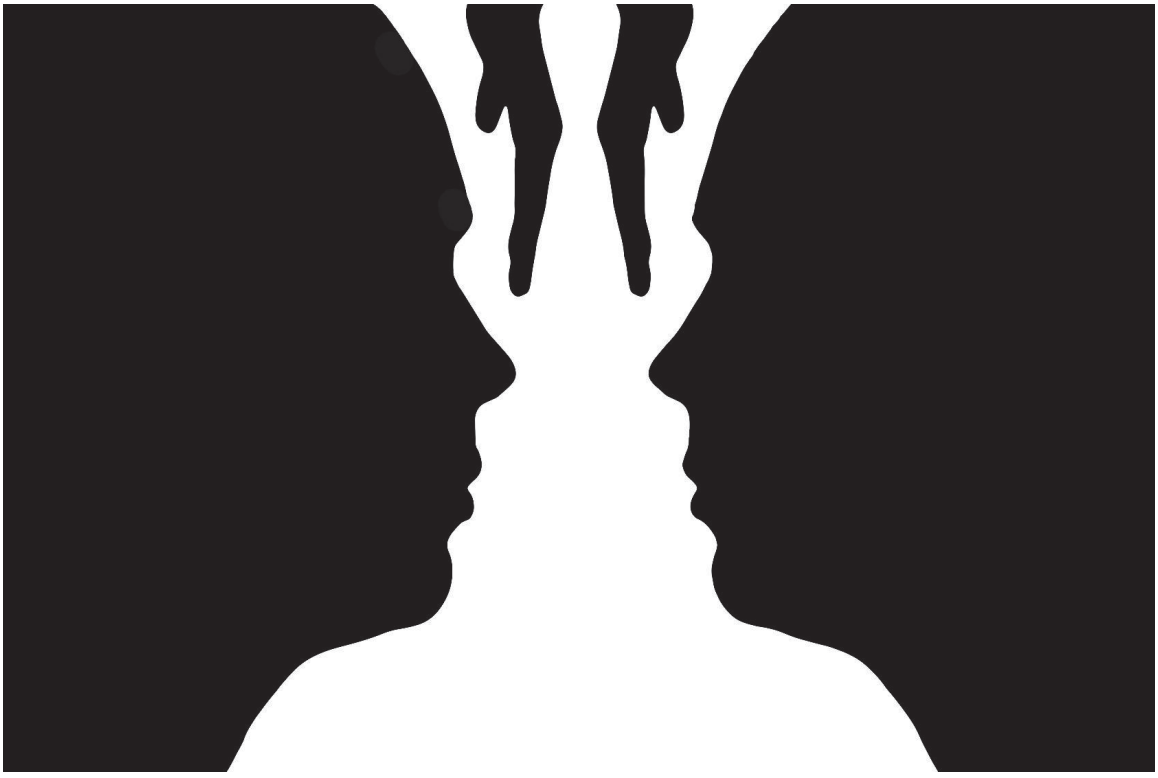
Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

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.

Download date: 07-11-2024

Chapter 8

Attempts to encourage diplomacy in online interactions: Three informative failures



Roos, C.A., Postmes, T., & Koudenburg, N (2022a). Attempts to encourage diplomacy in online interactions: Three informative failures. *Acta Psychologica*. Advance online publication. https://doi.org/10.1016/j.actpsy.2022.103661

We thank the ICT developers: Joost Timmerman, Wilmer Joling, and Robbert Prins for developing the chat environment used in Study 1. We additionally thank Pauliina Muikku for collecting the Study 1 data in the lab. Study 2 was made possible by Kira Choinski, who co-designed the experiment, developed the VoxBox, collected the data in the lab, and performed the exploratory content analysis. We are also grateful to dr. Mark Span for assisting in the development of the VoxBox. We thank Dona Beschka, Jan Failenschmid, and Lea-Sophie Steingruber for collecting the data of Study 3 and performing the content coding. We additionally thank Manuel Lentz for stepping in as a fourth coder.

The designs of studies 2 and 3 were preregistered prior to data-collection at osf.io/bkd7t and osf.io/7jfrc, respectively.

The data reported in this chapter are openly available in Dataverse at https://doi.org/10.34894/SAZIAE

Abstract

Online discussions about controversial topics seem more prone to misunderstanding and even polarization than similar discussions held face-to-face. Recent research uncovered an important reason why: certain behaviors that are used to communicate diplomacy and tact in face-to-face discussions – specifically, responsiveness and ambiguity – are more difficult to enact online. To improve online interaction experiences and understand the underlying mechanisms better, we ran three exploratory studies in which we tried to manipulate these diplomatic behaviors in online and face-to-face conversations.

Study 1 and 2 aimed to *increase* ambiguity and responsiveness in online environments to test whether it would result in increased experiences of solidarity. To this end, Study 1 (N = 68, repeated measures) compared a regular chat function with a chat function in which interaction partners saw each other's typing in real time. In Study 2 (N = 74, repeated measures), we introduced a keyboard that allowed participants to make interjecting sounds alongside text-based communication. In contrast, Study 3 (N = 105, repeated measures) aimed to *reduce* responsiveness and ambiguity in face-to-face discussion to test whether this would hamper participants' ability to navigate disagreements while maintaining solidarity. We asked participants about their conversational experiences both quantitatively and qualitatively in all studies.

We did not find the expected effects in any of the studies. The qualitative analyses of participants' behavior and commentary gave some insights into the reasons. Participants compensated for and/or distanced themselves from the manipulations. These behavioral adaptations all seemed to be socially motivated. We conclude by offering recommendations for research into online polarization.

Attempts to encourage diplomacy in online interactions: Three informative failures

There is increasing evidence that discussions via online media channels result in misunderstandings and even escalate into conflict and/or polarization more easily than face-to-face (FtF) discussions (e.g., Anderson et al., 2014; Coe et al., 2014; Settle, 2018; Yarchi et al., 2021). Research in this field has outlined how characteristics of online media, such as anonymity, a-synchronicity, or exposure to opposing views, affect *individual* attitudes and cognitions, such as perceptions of polarization (Yang et al., 2016), polarized attitudes (Bail et al., 2018), self-awareness (Nielsen, 2017), or disinhibition (Suler, 2004). A recent line of research took a different, complementary approach by studying *interpersonal dynamics* in interactions about potentially polarizing topics. This research demonstrated that certain behaviors that serve a diplomatic function in FtF discussions are more difficult to enact in text-based online discussions (Roos et al., 2020a, 2020b). Thus, while individuals' intentions may not differ, their behaviors tend to be less diplomatic online. Interaction partners, however, tend to misattribute this lack of diplomacy to each other's lack of social concern, which has consequences for their social relationship. Specifically, the reduced responsiveness (i.e., lack of instant feedback) online is interpreted as being less involved, and thus mistaken as a threat to the solidarity between interaction partners. And the reduced ambiguity or increased clarity online is interpreted as being more outspoken or extreme, which contributes to experiences of polarization between interaction partners. The current paper reports on three exploratory studies that tested interventions to increase these diplomatic behaviors in text-based online conversations in order to improve social outcomes.

Diplomacy Face-to-Face

Misunderstanding and polarization can occur both in FtF encounters and online. But whilst many studies have examined online polarization, far fewer have asked how people deal with the opinion differences they encounter in FtF conversations to prevent polarization. Recent research suggests that in FtF conversations, people infer the status of their relationship from the flow of their conversation. When conversation runs smoothly, interaction partners conclude that "things between them" are OK, whereas when the flow of conversation is disrupted by a frown or a silence, they may infer that something they said was problematic, and a potential cause for conflict (Koudenburg et al., 2011). Consequently, when the *content* of discussion is potentially contentious, people are well able to manage the relationship through the *form* of interaction. Indeed, maintaining a smooth

conversation in the face of disagreement suggests that this disagreement does not pose a threat to interaction partners' relationship, in other words, social harmony can be maintained (Koudenburg et al., 2017).

The pragmatics literature identifies two diplomatic behaviors that serve the flow and help maintain social harmony in potentially contentious FtF discussions: responsiveness and ambiguity. First, during FtF interaction, listeners continuously provide the speaker with feedback in the form of backchannel signals, which can be nonverbal, such as nodding, or verbal, such as interjecting sounds like "ah", "uh-huh", "hmm", "yes" (Beňuš, Gravano, & Hirschberg, 2011). This enables listeners to show the speaker that they are paying attention and encourage them to continue (Reis & Clark, 2013). This tends to make the speaker feel heard and gives them the impression that the listener is still "with" them, that is, understands what they are saying.

Second, when people in a FtF conversation express disagreement, or an opinion they expect might be controversial, they tend to *ambiguate* their language (Bavelas et al., 1990; Pomerantz, 1984). In such conflict-prone situations, people may prevent escalation by using disclaimers (e.g., "I do not know for sure"), hedges (e.g., "maybe," "sort of"), qualifiers (e.g., "some"), and hesitations (e.g., a drawn out "hmmm", "uhmmm's") (Brennan & Clark, 1996; Reid et al., 2003). These ambiguating cues are used to show consideration for the feelings of listeners that might disagree with the speaker. By signaling doubt or hesitancy about their position, the speaker also signals there is still room for discussion and persuasion. Thus, both responsiveness and ambiguity can be considered rather subtle diplomatic behaviors that enable participants in FtF discussion to maintain a sense of mutual understanding, or shared cognition, and good relationships, or solidarity.

Diplomacy Online

A recent line of research shows that the characteristics of text-based online media make it more difficult to enact these diplomatic behaviors (Roos et al., 2020a, 2020b). A limited synchronicity of communication makes interaction less responsive, and the text-restricted character of conversation makes expression less ambiguous, or clearer, (e.g., people express disagreement by stating "I disagree" rather than "I, ehm, maybe sort of do not totally agree", as they would FtF). Interestingly, whereas behavior changes when shifting from one medium to another, *expectations* concerning that behavior do not change. While people are likely unaware of these expectations, they do feel that their interaction partners respond differently online than FtF and try to explain this vague sense that something is off. But rather than recognizing that the relative lack of responsiveness and ambiguity might be due to the limitations the text-based medium poses on behavior, interaction partners tend

to misattribute this to each other's lack of social concern. Participants feel their partners disregard any diplomatic means because they are more concerned with clearly expressing their own opinion than with reacting to others. The resulting experience is that participants do not feel heard in such interactions (Roos et al., 2020a). This leads participants to conclude there is more disagreement between themselves and the others than there actually is; they see the conversation as polarized. Participants also experience less solidarity: they identify less with and feel more distant from their interaction partners (Roos et al., 2020a, 2020b).

Face-to-Face Diplomacy Online

A follow-up study explored the effects of inserting FtF responsiveness and ambiguity in online conversations by transcribing FtF discussions as if they were online chats and showing these together with original online chat discussions to a sample of observers (Roos et al., 2021). These observers were not aware of the origin of these transcripts and rated the degree of diplomacy they observed. They also evaluated the relationship between chatters. Results showed that, whereas responsiveness still served a diplomatic function in online discussions by increasing perceived shared cognition and solidarity, observers did not consider the original FtF discussions more responsive than the original chats. Moreover, whereas the original FtF discussions were seen as more ambiguous than the original chats, ambiguity had not such positive effects and even seemed to backfire by being perceived as evasiveness. This could be interpreted as evidence that what counts as diplomacy is medium-specific. However, there might be other, more medium-fit ways of manipulating diplomatic behavior that might be effective as de-polarizing mechanisms in online chats. Transcribing the FtF conversations resulted in a very awkward conversation with many incomprehensible sentences due to a lot of content-less utterances and unfinished arguments. Moreover, observers' evaluations of the solidarity experienced within an interaction may not accurately reflect how interaction partners themselves experienced their relationship.

Manipulating Diplomacy Online and Face-to-Face

The current paper reports on three exploratory studies in which we aimed to manipulate the degree of responsiveness and/or ambiguity within online or FtF discussions to change social perceptions and attributions and ultimately feelings of shared cognition and solidarity. The first two studies can be considered intervention studies where we tried to make online discussions more harmonious by encouraging interaction partners to be more responsive and/or ambiguous. The third study tried to do the reverse: make FtF discussions start less responsively and less ambiguously in order to test whether this threatens social harmony. In essence, in the first two studies we sought ways of making text-based chat more similar to FtF conversation,

and the third study worked the other way around, trying to introduce characteristics of text-based chat into FtF conversation.

In Study 1, we tried to increase online message ambiguity and responsiveness with a new chat function that showed participants what their interaction partner was typing in real time. We expected that, compared to a regular chat function where participants only saw "participant [1/2] is typing…", the insight into each other's message writing and editing would make participants perceive each other as more *in*hibited and less convinced of their opinion. This should not be dependent on the degree to which participants themselves feel inhibited and convinced. We also expected that participants might experience the conversation as higher in flow and feel more heard due to an increased responsiveness. These interpersonal perceptions (increased inhibition, reduced conviction), conversational experiences (increased flow), and feelings (feeling more heard) will increase the degree to which interaction partners experience shared cognition with, and solidarity towards each other.

In Study 2, we tested a keyboard, the "VoxBox", that allowed participants to make interjecting sounds (such as "hmhm", "aha") alongside their text-based conversation. We expected that, compared to a regular text-based chat, the addition of this VoxBox would increase responsiveness and thereby make both parties in the conversion feel more heard and experience more conversational flow, and consequently feel more solidarity towards each other.

In Study 3, we aimed to *decrease* responsiveness and ambiguity at the start of a FtF discussion by asking participants to read out their written down thoughts on the discussion topic. We expected that this manipulation would make the start of the FtF discussion more like a text-based chat discussion: with clear opinion statements and little responsiveness. This would lead participants to experience each other as more disinhibited, more convinced, and more extreme in their opinion, and the conversation as lower in flow. Hence, participants were expected to experience a reduced sense of shared cognition and solidarity.

Below we will describe all three studies in turn. This will be followed by a general discussion in which we synthesize the three studies and discuss their combined implications. All studies were approved by the Ethical Committee Psychology of the University of Groningen. Pre-participation informed consent was obtained from all participants in all studies. The designs of Studies 2 and 3 were preregistered prior to data-collection at https://osf.io/bkd7t/?view_only=a3cad5dc34384c74bb3ec99ac954c6cf and https://osf.io/7jfrc/?view_only=493453ced0f54f499972f25b858bf561, respectively.

Study 1

Background

We argue that an important reason for the higher ambiguity in FtF discussions results from people formulating their message while speaking, which introduces hesitations and hedges as people are searching for words. In essence, while FtF people observe each other's message editing, in the chat they only see end-products, which are rather clear. This suggest that giving chat participants insight into each other's message formation, might increase ambiguity, and thereby reduce perceived disinhibition and disagreement. Witnessing one's partner's edits might lead one to conclude that the partner is inhibited and takes one's feeling into account, and also that the partner is not that convinced about their opinion and might actually (partly) agree.

At the same time, seeing each other's message construction in real time might increase synchronicity and thereby responsiveness as participants can see that their partners start replying as soon as they themselves post something (or maybe even as soon as they start typing). One could even expect that a livelier chat would ensue where participants never complete or send their messages but interact in the message construction field, quickly and dynamically. This might promote the experience of conversational flow and make people feel more heard, which would also be profitable for their experienced agreement and solidarity.

Method

Research Design

The study was an experiment with a multilevel repeated measures design. Participants were part of a dyad. Each dyad participated in both the experimental and the control condition, in randomized order. Conversations in both conditions took place in a chat environment that was specifically designed for the current study. In the control condition, like in a normal chat, participants only saw "participant [1/2] is typing…" when their interaction partner was typing. In the experimental condition, participants could see *what* the other was typing when they were typing. Importantly, participants were aware that their partner could see their typing too.

Participant Characteristics

Participants were 68 native Dutch (M_{age} = 18.97, SD_{age} = 1.58; 82.4% female) students who participated for partial course credit. Most participants did not know each other before the experiment started (93%). The political orientation of the sample was right skewed with 58.8% placing themselves to the left, 25% placing

themselves in the middle, and 16.2% placing themselves to the right on a political orientation scale.

Power and Sample Size

Due to practical considerations, we were able to collect a sample of 68 participants, which provided a power of .80 to detect an effect of size $f = .17$ (small effect; Cohen, 1988) in a repeated measures within factors ANOVA with an alpha of .05.

Procedure

The study was conducted in Dutch. Participants were invited to the lab in duos. To ensure that they remained anonymous to each other, upon arrival, each participant was immediately seated into an individual cubicle with a computer. After both participants gave their informed consent, they received a list of 15 pre-selected statements that were dividing the Dutch politics at the time and were deemed relevant for students, for example: "The Dutch livestock must be halved" (see Table S1[50] for the full list of statements). Participants independently indicated whether they agreed or disagreed (dichotomous scale) with each statement. The experimenter then compared the ratings on all statements and marked two on which the duo disagreed (preferably choosing two statements where participants alternated their stances, e.g., participant 1 agreed with the first statement and participant 2 agreed with the second statement).

The participants then discussed the first statement for a maximum of 10 minutes either via a regular online chat (control condition) where the phrase "participant [1/2] is typing…" appeared in the chat screen when either of them was typing, or via the manipulated online chat (experimental condition) where each keystroke appeared in the chat screen immediately. All other chat characteristics were kept constant across conditions. After 10 minutes, the experimenter asked the participants to proceed to the questionnaire. When participants completed the questionnaire, the experimenter changed the condition by changing the settings of the chat environment on participants' computers. Participants then discussed the second statement in this second condition, again for about 10 minutes, and completed the questionnaire for a second time. In both chats, participants were pseudonymized as "participant 1" and "participant 2". All chats were screen-recorded and stored as transcripts. Finally, participants provided demographic details, were fully debriefed, and had the opportunity to ask the experimenter questions.

[50] Table numbers preceded by an "S" can be found in the supplementary materials.

Dependent Measures[51]

All dependent variables were measured on Likert scales ranging from 1 = *Strongly disagree* to 5 = *Strongly agree*. We measured perceived conversational flow with four items based on Koudenburg et al., (2017): "This conversation was …smooth/ …difficult (reverse coded)/ …pleasant/ …harmonious" (Omega ωh[52] = .65). Shared cognition was measured with three items adapted from Koudenburg et al. (2013a): "During this conversation, I felt that we …understood each other/ …were on the same wavelength / …agreed with each other" (ωh = .75). The experience of solidarity was measured with a three-item scale (Koudenburg et al., 2015): "During this conversation… I identified with the other/ …I experienced a sense of unity with the other/ …I felt that we were as one" (ωh = .84). We assessed to what extent participants felt heard by their interaction partner with two items: "During this conversation… I felt heard/ …I got the feeling that the other was ignoring me" (reverse coded) (ωh = .63). We measured the perceived inhibition of the other and of the self with single items: "During this conversation [the other/ I] thought carefully before saying anything." Perceived conviction of the other and conviction of the self were measured with three items: "[The other/ I] was …convinced/ …forceful/ …uncertain (reverse coded) about [their/my] opinion" (ωh = .63 for other and ωh = .70 for self).

Statistical Analysis

As participants were part of a dyad and were measured two times (in both conditions), the statistical analysis had to take into account these two sources of non-independence of observations. We therefore performed multilevel repeated measures regression analyses with condition (repeated measures; level 1), nested in participant (level 2), nested in dyad (level 3). We analyzed the data with the lmer function in the R package lme4 (version 1.1-21, Bates et al., 2019). For all dependent variables, we compared the fit of the multilevel repeated measures model that included only the random effect(s) of participant and/or dyad with the equivalent model that added condition as fixed effect predictor. We used the emmeans package (version 1.4.1, Lenth et al., 2019) to estimate condition means and confidence intervals.[53]

[51] The questionnaire contained a few more items that were discarded due to lacking reliability and/or relevance. Specifically, two additional inhibition items did not correlate well with each other (i.e., "the other could control themselves" and "the other got carried away"), and one item designed to measure the character of opinion expression by self and other did not correlate well with the other items (i.e., "[the other/I] was nuanced"). We also measured perceived (dis)agreement of the other and (dis)agreement of the self with the discussed statement, but this did not add relevant information to the current discussion.

[52] Hierarchical omegas with bias corrected and accelerated (1000) bootstraps. Calculated with the ci.reliability function of the MBESS package (version 4.6.0, Kelley, 2019).

[53] As there were no main effects for the order of conditions on the dependent variables, we did not include it as predictor in the models.

Results

As can be seen in Table 1, only one of the variables showed significant differences between conditions: the perceived conviction of the interaction partner. This effect was in the opposite direction of expected, however. In the experimental condition, participants felt that their interaction partner was *more* rather than less convinced of their opinion than in the control condition.

Table 1.

For each dependent variable, the f-test of the difference between conditions, and the means with 95% confidence intervals per condition.

Variable	*f*-value (1,67)	Control *M* [95%CI]	Experimental *M* [95%CI]
Conversational flow	0.30	3.95 [3.83, 4.08]	3.99 [3.86, 4.11]
Shared cognition	0.01	3.73 [3.50, 3.95]	3.74 [3.51, 3.96]
Solidarity	0.10	3.35 [3.16, 3.54]	3.32 [3.14, 3.51]
Feeling heard	0.41	4.33 [4.20, 4.46]	4.29 [4.16, 4.43]
Inhibition other	0.24	3.96 [3.76, 4.15]	3.91 [3.71, 4.11]
Inhibition self	3.76	4.19 [4.04, 4.34]	4.00 [3.85, 4.15]
Conviction other	5.32*	3.35 [3.20, 3.50]	3.57 [3.42, 3.72]
Conviction self	2.59	3.65 [3.48, 3.82]	3.50 [3.33, 3.66]

Note. * $p < .05$.

Qualitative Follow-up Analysis

To gain insight into the processes leading to the absence of expected effects, we analyzed participants' answers to an open-ended question that appeared towards the end of the questionnaire: "*Which differences did you experience between the two chats?"* We chose to perform an exploratory inductive thematic analysis since we wanted to uncover patterns of meaning – themes – in people's reported experiences

(Braun & Clarke, 2006). We also counted the number of participants mentioning each theme to have an indication of their (relative) prevalence.

Of the 68 participants, 17 only wrote about the factual manipulation (i.e., that they could see each other's typing in one condition and not in the other condition) or simply reported to what extent they agreed with their interaction partner. These were excluded from the thematic analysis, resulting in a sample of 51 answers. Three recurrent themes emerged: accelerated conversation, avoiding edits, and mind reading.

Accelerated Conversation. Many participants (45.10%) mentioned that the manipulation made the chatting quicker because they were able to start thinking about their response to their partner's message from the moment it was being constructed. Some participants experienced this speeding-up as pleasant, enabling more responsiveness.

"*In the 2nd chat, where one could see what someone was typing, I could respond faster because I had already read the other person's answer. Also, while the other person was typing, you could already think about what you were going to say and what your arguments are. The first* [control] *chat takes a little longer.*" (P. 50)

Other participants were not that positive and mentioned that the manipulation gave them less time to think. Interestingly, for some participants this led to more cross-talking in the experimental condition.

"*The first* [manipulated chat] *I tried to type in a hurry because we could both see what the other was writing. I wrote and read at the same time, which made it difficult in my head what I actually wanted to say. It was a bit like accidentally talking both at once in real life, both being quiet again, start talking again. That awkward moment, you know. The second* [control chat] *went much smoother because I now more often waited for a full answer.*" (P. 18)

Avoiding Edits. A consistently mentioned effect of the manipulation (33.33%) was feeling "watched" and therefore editing less. The following two quotes nicely illustrate participants' considerations.

"*That I could prepare what I was going to say in the second* [manipulated] *chat and I experienced the pressure of having to type it correctly in one go because the other was reading along.*" (P. 7)

"*In the second chat, I thought more about what I was typing because the other person could see it right away. I wasn't going to delete my sentence and type something else, something that I would do normally at times.*" (P. 62)

From these quotes it appears that participants prevented most of the in-text editing by thinking through their message *before* starting to type it. That is, the formulation of the message moved from the writing stage to the thinking stage.

Mind Reading. Interestingly, a couple of participants (11.76%) mentioned that they believed that the manipulation enabled them to follow the "train of thought" of their interaction partner. This suggests that these participants were not aware of their partner's behavioral adaptations to the manipulation (avoiding edits, theme 2), or that these participants' partners indeed did not adapt their behavior.

"that you can really see what someone types and sort of already see how the other thinks and reasons" (P. 44)

There was one participant that explicitly mentioned that their partner was *less* inhibited in the experimental condition.

"In the second [manipulated] *chat my interaction partner was more disinhibited."* (P. 56)

Discussion

The results of this study suggest that being able to see what one's interaction partner is typing in real time does not make one's perceptions of this partner more (or less) positive than in a regular chat where these message construction activities are hidden. The study's qualitative data offers a potential explanation for this by revealing that participants made behavioral adaptations to the manipulation that compensated its effects.

Some participants felt rushed because they tried to react to disappearing statements, which might have undermined the editing and the nuance in their messages, reducing ambiguity. This also led to more cross-talking, which undermines responsiveness. In FtF conversations this happens less often because there are many more cues helping participants to accurately predict who will be talking (e.g., Duncan, 1972). If cross-talking does happen there, it is quickly resolved by a hand gesture or a loud voice against a quiet one.

The qualitative data further suggests that participants compensated for the manipulation by editing less. Participants in the experimental condition stopped drafting their messages in the chat box and instead started drafting them in their minds; essentially hiding their thought process from their interaction partner. Participants reported they reduced their editing in order not to come across as doubtful or uncertain to their partner. What participants do not seem to be aware of, however, is that communicating doubt can be beneficial for their social relationship,

because their partner may perceive that they carefully (re)phrase their expressions to avoid potentially offending others (Roos et al., 2020a, 2020b).

Importantly, participants, while being aware of their own conscious censoring, believed that they got an accurate insight into each other's thought process. This, together with the reduced editing, might have led participants to perceive each other as slightly more rather than less convinced of their opinion in the experimental condition. This implies that the manipulation might have worked if participants were left unaware of the fact that their partner could see their typing, considering this would probably leave the editing intact and participants did tend to make internal attributions of the typing behaviors they observed.

In sum, we found no statistical support suggesting that the new communication mode we introduced improved the quality of social interaction. The qualitative analysis suggests this might be because the communication mode introduced various new disruptions and disturbances to the social interaction which we had not anticipated. Message senders showed compensatory behaviors, which made the manipulation less effective. Message receivers were unaware of this and instead interpreted the compensatory behaviors as indicative of the sender's true stance in the conversation and/or in the topic of discussion.

Study 2[54]

Background

Taking into account the limitations of Study 1, in Study 2 we offered participants the option to use a tool to "enrich" their text-chat conversation. Specially, we tested the effectiveness of a newly developed tool designed to make text-based chats more responsive. We expected that a heightened responsiveness would make interaction partners feel more heard and thereby experience more solidarity towards each other. The "VoxBox" is a small keyboard that complements the usual computer keyboard and produces interjecting sounds that are commonly used as backchannel signals in FtF conversation, such as "ah", "uh-huh", "hmm", "yes".

Method

Research Design

The present experimental study employs a repeated measures within-subject design with three conditions: Talking, Typing, and VoxBox. Participants took part in dyads and every dyad was exposed to all three conditions. The members of a dyad

[54] Part of the data from this study was presented in a paper on feeling heard (Roos et al., 2022b). The data for this study was collected in the context of a master's thesis (Choinski, K., 2020).

had a different role across all three conditions; an individual was either Sender or Responder. In all three conditions, the Sender was instructed to use a headset to tell the Responder about either their holidays, hobbies, or favorite tv-shows (all topics that are easy for most students to have a conversation about with another student they do not personally know). Therefore, the conditions were defined by the task of the Responder. In the Talking condition, the Responder was instructed to respond to the Sender via the headset. In the Typing condition, the Responder was instructed to respond to the Sender via a text-based chat function. In the VoxBox condition, the Responder was instructed to respond to the Sender via both typing in a text-based chat and the VoxBox. A Graeco-Latin square design was used to control for all possible order effects (of topic of conversation and condition) (Walker & Lev, 1953).

Development of the VoxBox

The VoxBox allowed Responders to make interjecting sounds as backchannel signals. Its content was based on a pilot study ($N = 6$) in which we carefully recorded all the interjecting sounds made by three dyads in three casual FtF conversations about the three topics used in the main study (one dyad per topic). The pilot participants were recruited from the peer group of the experimenter, a master student, and only informed about the exact research purpose during debrief. The experimenter instructed the pilot participants to "please talk about [the topic] for some minutes, as naturally as possible", after which she left the room. The three conversations were all audio-recorded and lasted for about five minutes. The recorded conversations were analyzed by the experimenter who coded and counted the interjecting sounds used as backchannel signals. These sounds were then sorted into groups based on their tone and meaning. For instance, "cool!" and "nice!" were considered the same group because of the similar tone and equal meaning (i.e., approving). This resulted in five frequently occurring sound groups (Table 2): approving vocalizations, laughing vocalization, neutral vocalizations, surprised vocalizations, and disapproving vocalizations.

The experimenter (female) reproduced and audio-recorded the individual sounds. A keyboard with five 1X1 centimeter buttons was built. The sounds were coupled with the buttons in OpenSesame (Mathôt et al., 2012). Each button represented one of the five groups that, when pressed, randomly produced one of the interjecting sounds from that group. The researchers pre-tested the VoxBox in conversation and made some last adjustments, such as reducing the delay between button-pressing and sound-playing. To help participants identify the right button, the buttons were labeled with either a symbol or a word representing the respective group. See Figure 1 for a picture of the VoxBox.

Table 2.
The five groups with their respective interjecting sounds. Sounds were grouped by tone and meaning. Symbols refer to the Jefferson Transcription System (n.d.)

Group	Sound
Approving	↓nice
	↓cool
	↓yes
Laughing	(h)a-
Neutral	Uh-hum
	Uhm
	Yeah
Surprised	↑oh:
	↑ah:
	↑Hm:
Disapproving	nah:

Figure 1.
The VoxBox.

Participant Characteristics

Participants were 74 students fluent in English (M_{age} = 20.18, SD_{age} = 2.24; 66.22% female, 31.08% male, 2.70% other), 72 of whom were first-year psychology students. They participated for partial course credit. 91.9% of the duos did not know or barely knew each other before the experiment started.

Power and Sample Size

A sensitivity analysis using G*Power (version 3.0.10, Faul et al., 2009) showed that this sample size yielded 80% power to detect between-condition (within-subjects) effects of at least f = .15 (small effect; Cohen, 1988), with alpha .05 (two-tailed), assuming sphericity and a correlation of .50 among repeated measures. Similarly, within-between interaction effects of at least f = .15 could be detected with 80% power, and alpha .05 (two-tailed).

Procedure

The study was conducted in English. Participants came to the lab in duos and were randomly assigned to the condition-order and topic-order of the Graeco-Latin square design. The experimenter asked the duos to split up in two individual cubicles. The cubicle that participants chose defined their role in the experiment (Sender or Responder) and came down to random allocation because participants were unaware of this fact. There was no visual contact: in none of the conditions could participants see each other. In their individual cubicles, participants read the instructions for their first condition on their computers. In the VoxBox condition, participants additionally received standardized oral instructions. Responders were also asked to try out the VoxBox to understand its workings and to get used to it. During this trial phase, the Senders were requested to put on the headset to also get acquainted with the tool.

After reading or hearing the instructions, the conversation started. Participants were allowed to end their conversation when it came to a hold. If not, the experimenter opened the cubicle doors after five minutes to end the conversation and ask participants to complete the questionnaire on their computers. This procedure was repeated for the other two conditions (with differing instructions depending on condition). After filling out the questionnaire for the third time, participants were asked to provide their demographics and were fully debriefed.

Dependent Measures[55]

All dependent variables were measured on Likert scales ranging from 1 = *Strongly disagree* to 5 = *Strongly agree*. First, feeling heard was assessed with the following four items of the Feeling Heard Scale (Roos et al., 2022b): "During this conversation …I felt heard by the other person/ …I could express myself freely/ …the other person respected what I [said/ wrote]/ …we misunderstood each other" (reverse coded) (ωh = .71). Secondly, as in Study 1, we measured conversational flow (Koudenburg et al., 2017) with four items: "This conversation was …smooth/ …difficult (reverse coded)/ …pleasant/ …harmonious" (ωh = .87). Third, we used four items adapted from Koudenburg et al., (2015) to assess solidarity, for example, "During this conversation …I felt a sense of belonging with the other person/ …I experienced a sense of unity with the other person/ …I identified with the other person/ …I felt accepted by the other person" (ωh = .82).

Statistical Analysis

Like Study 1, observations were non-independent (dyads, measured repeatedly). We again analyzed the data with the lmer function. For each variable, we defined multilevel repeated measures models with condition, role, and their interaction term as fixed-effect predictors, and participant nested in dyad as random effects. We again used the emmeans package to estimate condition means and confidence intervals. As our expectations concern the main effect of condition, we only report this effect here, and do not present the main effect for role nor the condition X role interaction effect.

Results

As can be seen in Table 3, there was a main effect for condition on all dependent variables. Across the board, Senders and Responders felt less heard and experienced less flow and less solidarity in the two online conditions than in the Talking condition. There were virtually no differences between the two online conditions. If there were, these were in the opposite direction of what was expected: participants experienced significantly less flow in the VoxBox condition than in the Typing condition ($t(144) = 2.425$, $p = 0.04$). This means that we did not find support for our expectations that the VoxBox would, by increasing responsiveness, make interaction partners feel more heard, experience more flow, and feel more solidarity.

[55] The questionnaire contained more dependent variables and single item measures than reported here (i.e., self-esteem, social presence, clarity of expression, politeness, disinhibition). These were discarded because they are not considered relevant to the current intervention aimed at increasing responsiveness.

Table 3.
Per dependent variable, the f-test of the main effects of condition, and the means and 95% confidence intervals in each condition.

Variable	*f*-value (2,144)	Talking *M* [95%CI]	Typing *M* [95%CI]	VoxBox *M* [95%CI]
Feeling Heard	45.75***	4.42_a [4.30, 4.54]	3.85_b [3.73, 3.98]	3.85_b [3.73, 3.98]
Conversational Flow	89.88***	4.34_a [4.16, 4.52]	3.32_b [3.14, 3.50]	3.07_c [2.89, 3.25]
Solidarity	19.65***	3.86_a [3.71, 4.01]	3.56_b [3.41, 3.71]	3.39_b [3.24, 3.54]

Note. *** $p < .001$. Means in the same row that do not share subscripts differ at $p < .05$.

Exploratory Qualitative Analysis

As the VoxBox was a newly developed tool and participants' experiences of it might help understand the quantitative results, we performed, like in Study 1, an exploratory inductive thematic analysis on the answers to the following open-ended question that appeared towards the end of the questionnaire: *"What did you think about the 'VoxBox' (the button box that made vocalization sounds)?"* The results of this analysis are outlined below.

42% of the sample evaluated the VoxBox negatively. The most frequently mentioned point of critique, by Senders and Responders alike, was that the sounds were too artificial. Senders felt like talking to a robot and mentioned it seemed like the Responder showed fake interest. Another point of critique, most mentioned by Responders, was that the device was too difficult to use in a conversation.

"I didnt like it. Her intonation was sarcastic and I didnt like talking to a bot" (P.6)

"I didnt like the sounds, they sounded very unrealistic to me." (P.57)

"This box made things more difficult than writing." (P.14)

Around 26% of the answers were predominantly but not entirely positive. These participants liked the idea and/or the device as such but had suggestions for improvement. The suggestions included that more "emotions" are needed (i.e., more buttons), that the VoxBox should be incorporated in the computer keyboard to make it easier to use, and that the sounds need to sound more natural. A male voice was also suggested.

"I think it is a good concept and the sounds it makes are sounds that happen during real conversations, but the voice just sounded too artificial." (P.2)

"Kinda ok, but still needs a lot of improvement. More options, a man's voice, a much more natural voice." (P.34)

17% of the participants liked the VoxBox, apparently unconditionally. These participants stated that the VoxBox improved the conversation and made it livelier by reducing the restrictions of conventional typing and adding a social aspect.

"It was nice for an immediate feedback without waiting for the typed message." (P.38)

"I liked it a lot! It was nice to hear a laugh or 'mhhh' every now and then in the conversation, you do feel like the other person understands you more." (P.60)

The remaining participants (15%) gave a neutral response, stating that the VoxBox did not add anything to the conversation but also did not disturb it.

"It didn't really bring much to the conversation. Saying "ah" doesn't save the time at all." (P.12)

Discussion

The VoxBox did not appear to make online reactions more responsive. It did not make interaction partners feel more heard or experience more solidarity. Participants even experienced their conversation as flowing more smoothly when they only used the text-based chat function than when they were also allowed to use the VoxBox. The open-ended answers gave some insight into why this might be the case: the use of the VoxBox made conversation feel unnatural, artificial, and even insincere. This suggests that responsiveness is something very delicate and precise and therefore hard to manipulate.

Thus, similar to Study 1, we find that the intervention did not produce the improved quality conversation we had hoped it would. The qualitative analysis suggests that, just like in Study 1, the novel features of this condition introduced new disturbances in the social interaction which, at least for some participants, made the experience worse rather than better.

Study 3

Background

We tried to improve online conversation in Study 1 and Study 2 by enabling the use of face-to-face diplomatic behaviors. Both interventions did not work and this raises the question whether the cross-medium transplantation of diplomatic behaviors is feasible at all. The current study therefore takes the reverse approach by disabling these diplomatic cues in FtF discussions. By essentially mimicking some of the characteristics of online chats in FtF conversation we aimed to gain more insight

into the processes enabling responsiveness and ambiguity and their positive social outcomes in FtF discussions.

We assume that the degree of ambiguity and responsiveness peaks at the start of a FtF discussion about a controversial issue among strangers as participants are aware that the topic potentially elicits disagreement but are still unaware of each other's opinions and therefore first "test the waters". As such, the difference in ambiguity and responsiveness between chat and FtF could be most pronounced at the start of discussions; the point where it is also most consequential for the subsequent evolution of the discussion and its outcomes. Indeed, blatant opinion differences combined with the perception of strong convictions of participants at the start of online discussions could discourage their attempts to come to an agreement. This study therefore aimed to isolate the effect of opinion clarity and conversational unresponsiveness at the start of a FtF discussion.

Method

Stimulus Materials

We conducted a small pilot study to select topics with a high potential for instigating a controversial discussion among participants. The pilot sample consisted of fourteen first-year psychology students (M_{age} = 20.79, SD_{age} = 2.49; 50% female) that received partial course credit. These participants were recruited from the same pool as the intended sample for the main study. Political orientation was right skewed: 78.6% placed themselves on the left, 14.3% placed themselves in the middle, and 7.1% placed themselves to the right on the political spectrum.

We designed an online questionnaire that included 30 controversial statements we deemed relevant to students. For each statement, pilot participants were asked to indicate how strongly they felt about the topic, whether they thought other students would agree with them (reverse coded), the degree to which they considered the topic personally relevant to themselves, and the extent to which they considered themselves able to have a five-minute discussion about the topic. The ten statements that scored highest on the combined score were selected as stimulus material for the main study, for example: "Companies should be required to hire 50% male and 50% female employees" (see Table S2 for the full list of statements).

Research Design

The main study was an experiment with a multilevel (individuals nested in groups) repeated measures (of condition) design. Groups of three students participated in both conditions in randomized order. In both conditions, participants discussed about a topic on which they disagreed. In the control condition they received no instructions on *how* to discuss the topic, such that a natural discussion would emerge.

In the experimental condition participants first wrote down and then read out their thoughts on the topic at the start of the discussion.

Power and Sample Size

Our full design required 33 triads, which came down to a sample size of 99 participants. A power analyses using G*Power (version 3.0.10, Faul et al., 2009) confirmed that the projected sample size would result in adequate power at $p = .05$ and effect size of $f = .25$ (medium effect; Cohen, 1988): .91 at the group-level (i.e., if group-level ICC's were 1) and .999 at the individual participant level (i.e., if group-level ICC's were 0).

Participant Characteristics

Participants were 105 students ($M_{age} = 21.40$, $SD_{age} = 3.22$; 63% female) who participated either for partial course credit or monetary compensation. The vast majority (96.5%) indicated to barely know or not at all know their group members prior to the experiment. Their self-reported political orientations were 74.3% left-wing and 13.3% right-wing; 12.4 placed themselves in the middle of the political spectrum. All participants considered themselves proficient in English.

Procedure

The study was conducted in English. Participants were invited into the lab in triads where they were immediately seated in separate cubicles behind a computer. After reading the study information, the experimenter provided each of them with a list of the ten controversial statements selected in the pilot study. Participants were asked to indicate whether they agreed or disagreed (dichotomous scale) with each statement. When all participants were done, the experimenter compared their ratings and marked two statements on which the group members had diverging opinions (we tried to alternate the topics as much as possible between groups). Participants were given back their rated statement list and told they would discuss the marked statements, starting with the first, because there was disagreement about these statements in their group. The following steps of the experiment were repeated two times: once in each condition.

In the experimental condition, participants were instructed to *"Write down your thoughts on this topic"* on a piece of paper. In the control condition, participants were instructed to *"Think about this topic"*. Participants could always take as much time for this as they needed. When they felt they were ready, participants entered an adjacent room where they were seated on chairs in a triangle. In the experimental condition, participants started the discussion about the topic by reading out their written down thoughts to each other in turn, after which they were free to discuss the topic further. In the control condition, participants started a natural FtF conversation and discussed the topic freely from the start. After discussing for five minutes or less (participants were

allowed to end the discussion within this time frame), participants returned to their cubicles to fill out the questionnaire on their computers. To enable content coding, all conversations were audio-recorded.

Finally, after the second condition, participants provided their demographic details and were given the opportunity to ask the experimenter questions. The full debrief was emailed to all participants after the data collection was complete.

Dependent Measures[56]

Questionnaire. Participants rated all items on scales ranging from 1 = *completely disagree* to 5 = *completely agree*. Like studies 1 and 2, we measured perceived conversational flow (Koudenburg et al., 2017) with four items: "This conversation was …smooth /…difficult (reverse coded)/ …pleasant/ …harmonious" (ωh = .69). To measure perceived shared cognition, participants rated four items adapted from Koudenburg et al. (2013a): "During this conversation I felt that the other participants and I …understood each other/ …agreed with each other/ …were on the same wavelength /…had a different opinion (reverse coded)" (ωh = .75). Solidarity was assessed with three items from the belongingness subscale of the need threat scale (Van Beest & Williams, 2006), for example: "During this conversation, I felt I belonged", combined with the single item social identification measure of Postmes et al. (2013): "During this conversation, I identified with the other participants", and a self-devised item: "During this conversation, I had a good relationship with the other participants" (ωh = .81). Similar to Study 1, we measured the perceived inhibition of the others and of the self with single items: "During this conversation [the other participants/ I] thought carefully about how [they/I] expressed [themselves/myself]." Perceived conviction of the others and conviction of the self were both measured with one item: "[The other participants/ I] [could/can] easily be convinced to change [their/my] opinion" (reverse coded). Perceived opinion extremity of the others was also measured with a single item: "The other participants had an extreme opinion".

Discussion Coding. To check the effectiveness of our manipulation, we coded the first three turns, the turns that were read out by participants in the experimental condition, of each conversation on clarity (1 = *Very ambiguous*, 2 = *Ambiguous*, 3 = *Neutral*, 4 = *Clear,* 5 = *Very clear*; ICC = .51) and responsiveness (0

[56] We measured additional codes and single items which are not reported due to lacking reliability, missing observations, and/or lacking relevance to the current discussion. The additional codes were: expressed disagreement, disinhibited behavior, and the number of arguments or reasons the speaker provided in support for or against their opinion. The additional items were: participants' own and perceived other's opinion on the discussed topic, own and other's certainty of opinion, own and other's politeness of expression, own and other's considering consequences, own and other's tentativeness of expression, own and other's clarity of expression, own and other's true opinion expression, and other's nuanced opinion.

= *No*, 1 = *A bit*, 2 = *Yes*, ICC = .84). This coding scheme was based on the schemes used in Roos et al. (2020a, 2020b). Generally, the more and the stronger the expressed ambivalence, disclaimers, and hedges (e.g., "I don't know for sure," "as far as I know," "sort of"), the more ambiguous a statement was considered. When participants presented their opinion as a fact, this was rated as very clear. Responsiveness indicated for each turn whether it connected to the turn directly preceding it. When a turn started with a connecting word (e.g., "yes," "no," "but") and contained a reaction to the preceding turn, we coded it as responsive. When a connecting word was missing but the previous speaker was acknowledged, we coded it as a bit responsive. When the previous speaker's turn was ignored, this was classified as unresponsive. The very first turn of a conversation was not coded for responsiveness as there was no preceding turn.

All conversations were double-coded. Four trained assistants independently coded half the discussions. Of the coders, three were familiar with the research and its aims and one was blind. Coders coded the untranscribed audio-recordings to retain more of the interactions' character. To assess the inter-rater reliability of these ordinal codes, we calculated two-way absolute agreement average measures intra-class correlation coefficients (Hallgren, 2012) with the icc function in the irr package in R (version 0.84.1, Gamer et al., 2019). We first calculated the means per conversation per coder (averaging the scores on the three turns), as we wanted to analyze the data at this level. The codes that were of insufficient reliability, were partly recoded by the coders collectively. We took the average of both coders' scores as input for the analysis.

Statistical Analysis

As in Study 1 and 2, we performed multilevel repeated measures regression analyses. For all dependent variables and codes, we compared the fit of the multilevel repeated measures model that included only the random effect(s) of participant and/or triad with the equivalent model that added condition as a fixed effect predictor. We again analyzed the data with the lmer function and used the emmeans package to estimate condition means and confidence intervals.

Results

As can be seen in Table 4, we did not find any differences between the experimental and control condition in participants' subjective experiences of the conversations. Participants in both conditions experienced equal amounts of conversational flow, shared cognition, and solidarity, and considered their partners and/or themselves equally inhibited, convinced, and extreme in opinion.

Table 4.
For each dependent variable, the f-test of the difference between conditions, and the means with 95% confidence intervals per condition.

Variable	*f*-value (df)	Experimental *M* [95%CI]	Control *M* [95%CI]
Flow	0.294	4.21	4.26
	(1,104)	[4.04, 4.38]	[4.09, 4.43]
Shared Cognition	0.015	3.90	3.91
	(1,174)	[3.69, 4.12]	[3.70, 4.13]
Solidarity	0.114	4.19	4.17
	(1,104)	[4.01, 4.38]	[3.99, 4.35]
Inhibition other	0.60	4.45	4.51
	(1,104)	[4.29, 4.60]	[4.35, 4.66]
Inhibition self	0.00	4.33	4.33
	(1,104)	[4.11, 4.56]	[4.11, 4.56]
Conviction other	0.26	3.24	3.31
	(1,104)	[2.97, 3.50]	[3.04, 3.57]
Conviction self	2.28	3.67	3.50
	(1,104)	[3.39, 3.95]	[3.22, 3.78]
Extremity other	0.035	1.64	1.62
	(1,104)	[1.41, 1.87]	[1.39, 1.85]

Note. None of the *f*-values reported in this table is significant at $p < .05$.

This might be explained by the content analysis showing that the manipulation was only partially effective (see Table 5): whereas the read-aloud-turns in the experimental condition were less responsive than the first three turns in the control condition, there was no effect on clarity. The coders reported two possible reasons for the latter: participants tried to "soften" the impact of their statements by a) using ambiguating techniques and/or b) demonstratively showing that they were doing the assignment. First, participants ambiguated their written down statement upon reading it aloud by embedding it in a lot of qualifiers (e.g., "a bit") and hedges (e.g., "ehm"), like people tend to do in contentious FtF discussions. Second, participants communicated to each other that the way they were acting was "just" due to the experimental instructions. They did so explicitly by preceding their statements with phrases like "I wrote the following…" or more implicitly by mumbling or reading fast as if to show they did not really mean what they said.

Table 5.
For both codes, the f-test of the difference between conditions, and the means with 95% confidence intervals per condition.

Variable	*f*-value (df)	Experimental *M* [95%CI]	Control *M* [95%CI]
Clarity	0.830 (1,68)	4.00 [3.86, 4.14]	3.91 [3.77, 4.05]
Responsiveness	24.30*** (1,34)	2.95 [2.74, 3.16]	3.59 [3.37, 3.80]

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

Table 6 shows the correlations between coding and dependent variables. Only the correlation between clarity and the mean perceived conviction of interaction partners was significant and of medium effect size (Cohen, 1988). This correlation suggests that participants in FtF conversations characterized by clearer expressions at the start of the discussion perceive each other (but not themselves) as more convinced of their opinion. This effect is in line with our theorizing but might be due to chance since the other correlations between coding and dependent variables, including the variables that should be strongly related to conviction, were not significant. This suggests an alternative explanation for the null effects, one that is not in line with our theorizing: the degree of clarity and responsiveness in the first three turns of a conversation does *not* affect interpersonal perceptions in this FtF situation.

Table 6.
The repeated measures correlations between the codes (columns) and the dependent variables (rows) at the group-level.

Variable	Clarity	Responsiveness
Flow	.16	.01
Shared cognition	.10	.09
Solidarity	-.13	.01
Inhibition other	.27	.13
Inhibition self	.01	-.03
Conviction other	.34*	-.11
Conviction self	.25	-.09
Extremity other	-.12	-.09

Note. * $p < .05$.

Discussion

Reading aloud their written opinion in turns at the start of a FtF discussion, did not damage participants' sense of conversational flow nor their experiences of shared cognition and solidarity towards each other. Like in Study 1 and 2, the qualitative analysis shows that people adapt their behavior to the new possibilities and restrictions offered by the new mode of interaction we created for them, in such a way that they buffer the negative effects of the manipulation. How exactly people adapt is very informative.

In essence, participants in this study translated their text-based statement to something they deemed fitting and appropriate in a divided FtF discussion with strangers. In a mirror image of Study 1, where chatters tried to write clear and unambiguous statements when we tried to make their uncertainties transparent, participants in the FtF discussions in the current study invested in making their message more ambiguous when we forced them to be clear. They did so either by ambiguating their statement with qualifiers and hedges, or by renouncing the statement as due to the assignment and not being an accurate reflection of their true opinion. Thereby, participants signal hesitancy and doubt when sharing their opinion. In this way, participants distance themselves from the clear statement they wrote down.

It is also interesting to note what participants did *not* do: they did not make their statements more responsive when reading them aloud (i.e., the first three comments were significantly less responsive in the experimental condition than in the control condition). Participants in the experimental condition thus did not insert connecting words at the start of their comment and/or explicitly acknowledge the previous speaker's comment. This might be due to participants not noticing this reduced responsiveness and/or not recognizing it as a potential threat to solidarity. In line with the latter, the reduced responsiveness in the first three statements seems to have had no negative consequences for the relational outcomes; indeed, none of the dependent variables significantly correlated with responsiveness. This lack of negative consequences might result from people being explicit about behaving in line with the experimental instructions and thereby pushing the blame for the reduced responsiveness away from themselves.

General Discussion

In three exploratory studies, we introduced three different interventions to manipulate responsiveness and ambiguity. In the first two studies we tried to promote them, and in the third we tried the reverse. None of these interventions successfully affected responsiveness and ambiguity, however. The qualitative

findings shed light on why this might be. The interventions introduced novelties into social interactions that participants were unfamiliar and uncomfortable with. Essentially, they worked around the intended behavioral changes, by finding creative new ways of keeping their old behavioral patterns for each medium intact. The results show that whilst it is possible to intervene in very well-learned patterns of (online and FtF) behavior, people readjust their behavior to counteract the effects of the intervention.

Behavioral Responses

We gained insight in participants' behavioral responses through questioning them after each experiment in studies 1 and 2 and through coding their communication in Study 3. In Study 1, participants reported that they adapted their behavior, and this may explain why the manipulation had no effect. Specifically, they tried to avoid their partner seeing them formulate and edit their message in the text-based environment. Participants first thought and then typed their message rapidly without correcting themselves. So, they "edited out" their uncertainty and hesitancy. In Study 2, where the Voxbox introduced a new layer of interaction that they could not ignore or undo, participants found the Voxbox intrusive: it seemed to violate their expectations for how chats should go, and recipients appear to have "edited out" the intended social signals. Participants experienced difficulties using the VoxBox in their interaction and felt the resulting communication was insincere and disrupted the conversational flow. In Study 3, participants adapted their behavior, undoing the manipulation. Specifically, participants did not simply read aloud their opinion but softened it up and ambiguated. So, they "edited in" their uncertainty and hesitancy. In this set of results, two patterns stand out: 1) compensating for the intervention and 2) distancing from the intervention.

Compensating for the Intervention

First, in studies 1 and 3, participants adjusted their behavior to undo the intended behavioral effects of the interventions, and the way they did provides interesting insights into how behavioral norms differ per medium. In the online context of Study 1, people tried to be clear and in the FtF context of Study 3, people tried to be ambiguous. This finding is in line with recent research showing that, whereas ambiguity serves a central function in socially regulating FtF discussions, it can be considered indicative of conflict when perceived in a text-based chat environment (Roos et al., 2021). These findings can be explained by expectancy violations theory (EVT, Burgoon and Hale, 1988; Burgoon and Walther, 1990). EVT proposes that people start an interaction with certain assumptions about the behavior that is accepted and expected in that context, that is, the behavior that is normative. These assumptions are based on people's prior experiences in similar

interaction contexts, such that, for example, one comes to expect a certain friend to be quite outspoken, or such that one expects online business calls to be relatively disorganized. In a similar way, norms specifying which diplomatic behaviors are appropriate may differ between media, and the current studies show that people try to conform to these norms. In essence, the manipulations in studies 1 and 3 tried to steer participants' behavior in anti-normative directions by encouraging ambiguity in an online context and clarity in a FtF context. But participants creatively worked around these manipulations to stick to the behavioral norms that are "fitting" with the communication context. This implies that diplomatic cues are medium-specific and should be analyzed in the specific communication context in which they exist.

These norms appear to be related to, maybe even based on, the behavioral possibilities the medium affords. When people discuss some sensitive topic with strangers FtF, they tend to continually re-construct and edit their statements while speaking in response to the (often subtle) cues of (dis)approval their interaction partners emit (Pomerantz, 1984). In essence, people probe what they can and cannot say by speaking tentatively. Study 3 shows just how strong this tendency is: even when receiving clear instructions to read out aloud their carefully formulated written statement, participants tend to edit in all sorts of cues that show hesitancy and uncertainty during their vocalization. These regulatory cues are not or less available in text-based online conversations (Roos et al., 2020a, 2020b). It therefore seems reasonable that participants in Study 1 preferred to think through and carefully formulate their message before showing any of it to, and potentially offending, the other. It thus seems that, in both cases, communicators tried to be diplomatic, but in a way that fits the medium's possibilities and norms.

Distancing from the Intervention

The second pattern that emerges concerns studies 2 and 3 where participants experienced the changes to their regular communication routines in a certain medium as intrusive and complicated to deal with. In Study 2, participants reported this after the experiment: using the VoxBox felt unnatural and came across as insincere. Although we tried to make the sound-response of the buttons as instantaneous as possible, there was still some time delay, increased by participants having to seek out the right button. Participants experienced this as disrupting the natural flow of their conversation: participants experienced least flow in the VoxBox condition, compared to when they could interact via headphones or text-chat only. Previous research in the context of FtF interaction has shown that while deviations from "normal" interaction can be subtle, they may be experienced as disruptions of conversational flow, and as such, have serious implications for relationships by raising questions about the solidity of relationships and mutual understandings

(Koudenburg et al., 2013b, 2017). Similarly, research in the domain of online communication has shown that online communicators easily notice response latencies and that this affects their perceived relationship (Walther & Tidwell, 1995). Indeed, in Study 2 we also find the lowest levels of solidarity in the VoxBox condition where response latencies were most unnatural. In Study 3, participants seemed to prevent such negative relational consequences by distancing themselves from the experimental intrusion in the interaction itself. For instance, they did so by communicating that they were "just" following the experimental instructions, either quite explicitly, by saying things like "So, I've written down…", or more implicitly, by speaking monotonously when reading out what they had written, in a way distancing themselves from their stated opinion. People thus seem to create an external attribution for the disrupted flow: "It is not me, it is the experiment". This can be considered a face-saving strategy (Brown & Levinson, 1987). In general, studies 2 and 3 suggest that deviations from well-learned behaviors in communication contexts are experienced as unnatural and unpleasant.

In sum, in all studies, the interventions introduced unfamiliar aspects into well-known communication contexts, and these changes made participants feel uncomfortable. Participants were eager to compensate for (studies 1 and 2) or distance themselves from (studies 2 and 3) the manipulations. This suggests that participants experienced the manipulations as disrupting their conversation and as a threat to the relationship with their interaction partner(s). It also suggests that participants (consciously or not) were motivated to preserve a pleasant conversation and a good relationship, and that they tried to accomplish this by falling back on the behavioral norms they associate with the medium.

Relational Consequences

Although the present research suggests that participants counteract intrusions to conform to medium-specific communication norms, the question is whether these behavioral adaptations may always serve the relationship. Whereas it might be written thoughtfully and carefully, a clear online message could easily be misinterpreted and mistaken for being too blunt and outspoken (see also Roos et al., 2020a, 2020b). Indeed, interaction partners might see the well-formulated self-contained online statements as a sign the sender is more concerned with venting their own opinion, of which they are strongly convinced, than with listening to them. Interaction partners can therefore feel unacknowledged and not heard (Roos et al., 2022b). This is quite paradoxical: whereas online communicators are socially concerned and try to be diplomatic by formulating as precisely as possible, they come across as self-focused and non-diplomatic by doing so. This was also shown in Study 1, where participants felt that their partner was quite convinced of their opinion

because the partner typed in one go. This is a clear case of misperception: participants thought they had an insight into their partner's real thought process, which thus seemed to be quite self-focused, while reporting they themselves concealed their thought process from their partner (by thinking before typing), which was in that context essentially an other-focused act. So, whereas people try to be diplomatic and considerate of others in the online context by being precise and clear, they take the same behaviors performed by others as a sign of missing diplomacy and lacking social concern. This seems to be the fundamental attribution error at work (Jones & Harris, 1967): whereas participants acknowledge that the technology influences their own behavior (external attribution), they think the same behavior from their partner is due to malicious motivations (internal attributions). This tendency for internal misattribution seems to be largely unconscious and hard to change: even when communicators know their partner's behavior is due to the limitations of the medium, they still attribute their partner's behavior to a lack of social concern (Koudenburg et al., 2013b).

Limitations and Future Directions

The conclusions in all three studies result from exploratory qualitative analyses and should therefore be considered as observation-based hypotheses (input for theory formation) rather than as confirmations of effects predicted from theory. This means that quantitative research is required to confirm these ideas.

Part of the explanation of why manipulations were not effective is that participants were simply not used to the communication context and therefore anchored their expectations and their behavior to the most similar communication context that *is* familiar to them (i.e., text-based chats in Study 1 and 2, and FtF conversation in Study 3). This raises the question of whether our manipulations would have been more effective when people would get used to the new communication context, such that their expectations would be adjusted accordingly. This is left for future research to establish.

Conclusions

It is often assumed that online discussions are prone to polarization because people are less socially concerned and more disinhibited when they go online (Nielsen, 2017; Suler, 2004). In contrast, the patterns of behavioral adaptations observed in this set of studies suggest that people continue to be very socially concerned online, but their attempts may backfire. Specifically, when we introduced interventions that we anticipated would promote diplomatic behaviors, participants sometimes resisted by adjusting their actions, believing that to be best for the interaction and their social impression. Participants felt the interventions disrupted

the interaction and tried to repair or prevent relational damage by formulating themselves as precisely as possible. But paradoxically, the resulting comments are so clear and self-contained that interaction partners conclude the sender is very convinced of their opinion and not concerned with maintaining a pleasant conversation and a good relationship.

This set of studies therefore suggests that the best way to promote harmony and prevent polarization online may not be to change behavior (because people will push back and inadvertent negative effects will occur) or to change motivations (because people are already socially motivated), but rather to improve dialogue. This involves encouraging a mutual understanding between interaction partners, as well as preventing misinterpretation and misattributions of each other's behaviors, all within a certain communication context. This implies that to understand and intervene in online polarization, we need to shift focus from the individuals, and their motivations and behaviors in the interaction, to them as part of an interacting social unit, in which behaviors and their interpretations are shaped by the specific communication context.

Supplementary Materials

Table S1.

The list of 15 statements used in Study 1.

No.	Statement
1	Teachers should always decline friend requests of students on social media.
2	Flight tickets must be made twice as expensive worldwide to save the environment.
3	Smokers and drinkers should be at the bottom of waiting lists for organ transplants.
4	The Dutch livestock must be halved.
5	People who left to Syria are no longer allowed to return to the Netherlands.
6	Artificial intelligence poses a threat to humanity.
7	Everyone should be able to say what they want without being accused of inciting hatred.
8	Setting off fireworks yourself is a tradition that must be preserved.
9	The government should make children's vaccinations mandatory.
10	Everyone should be able to retire after 45 years of work.
11	You are allowed to film violent crimes on the street and share this on social media.
12	Dutch identity is threatened by all immigration.
13	All pedophiles should be locked up.
14	It's time for a women's quota in Dutch industry.
15	Euthanasia should be made possible for children younger than 12.

Table S2.

The list of 10 statements used in Study 3.

No.	Statement
1	Companies should be required to hire 50% male and 50% female employees.
2	Meat should be more expensive.
3	Social media benefits people's social lives.
4	Voting should be made mandatory for all citizens.
5	Drug use should be treated as a mental health issue rather than a criminal offense.
6	Student loans are exploitative.
7	All pedophiles should be locked up.
8	It should be allowed to film harassments in the street and share them on social media.
9	Privacy is less important than safety.
10	Homework is necessary for learning.