Slowing it Down: Towards Facilitating Interpersonal Mindfulness in Online Polarizing Conversations Over Social Media

TEALE W. MASRANI, University of Calgary, Canada
JACK JAMIESON, NTT Communication Science Labs, Japan
NAOMI YAMASHITA, NTT Communication Science Labs, Japan
HELEN AI HE, University of Calgary, Canada

Discussions about polarizing topics are essential to have, yet they can easily become hostile, aggressive, or distressing on current social media platforms. Content moderation interventions aim to mitigate this issue, though such approaches are reactionary, removing harmful content only after it has been posted. We conducted a mixed-methods experiment with 40 participants to investigate how a design friction that manipulates the temporal flow *during* a contentious conversation can foster interpersonal mindfulness, a trait critical for productive communication. Dyads were randomly assigned into the Control Group which received no intervention, and the Experiment Group where participants were limited to sending one message per two-minute interval. Triangulating quantitative and qualitative data from conversation logs, questionnaires, interviews, and computational text analysis, our findings revealed a two-fold effect: Experiment Group participants felt simultaneously frustrated by the intervention as it disrupted the pacing of their conversation and interfered with rapport-building, and appreciative of the intervention as it nudged them towards writing more thoughtful and task-focused messages. We discuss implications of these findings for future investigation into the design of temporal interventions to influence interpersonal mindfulness during contentious online discussions.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: social media, design frictions, polarizing conversations, interpersonal mindfulness

1 INTRODUCTION

Discussions about contentious or polarizing topics can bridge ideological gaps between people with dissenting perspectives, promote a greater understanding of complex issues, and create cohesion and inclusion within a heterogeneous society [23, 53, 78]. Social media platforms such as Facebook, Reddit, and Twitter offer an ideal space to have these conversations by bringing individuals together from diverse geographical regions with varying belief systems. Indeed, studies find that people often

Authors' addresses: Teale W. Masrani, tealemasrani@gmail.com, University of Calgary, Canada; Jack Jamieson, jack@jackjamieson.net, NTT Communication Science Labs, Japan; Naomi Yamashita, naomiy@acm.org, NTT Communication Science Labs, Japan; Helen Ai He, helenaihe.jobs@gmail.com, University of Calgary, Canada.

discuss contentious topics on social media [3, 7]. Yet, such discussions can easily become distressing or unproductive, often involving aggressive and hurtful behaviour [7, 14, 44, 51]. Consequently, some social media users actively avoid engaging in these conversations, due to the possible emotional toll of encountering disagreement or damaging relationships [7, 78]. To alleviate this, current efforts have focused on human or automated content-moderation – that is, removing harmful or toxic user content, or restricting what users may post. However, such interventions can be ineffective as they: i) disproportionately affect already marginalized users [31], ii) can be perceived by individuals as unfair [57], and iii) can hinder user experience, causing users to migrate to niche social media platforms with less restrictions [82].

As an alternative to content-moderation approaches that remove harmful content *after* it has been posted, this paper explores the potential for a design intervention to minimize reactivity and verbal aggression *during* contentious conversations. Specifically, we investigate how manipulating the temporal flow of a conversation may influence conversational dynamics and ultimately support interpersonal mindfulness – a trait critical for productive communication [17, 18, 45]. We present the results of an exploratory, mixed-methods experiment where 40 participants (20 dyads) engaged in a synchronous, text-based, negotiation task about salient polarizing issues. Ten dyads were randomly assigned to the Control Group, which received no intervention, and ten dyads to the Experiment Group, where participants could send one message per two-minute interval.

We triangulated quantitative and qualitative data from conversation logs, questionnaires, interviews, and computational text analysis, and our findings revealed two main dichotomies in the influence of the intervention. First, while Experiment Group participants were frustrated by the two-minute delay, they appreciated how the delay encouraged more thoughtful message construction. Second, while the intervention interfered with the ability to build rapport, Experiment Group participants appreciated a greater focus on the goal of the conversation. We cannot conclude that the intervention had an overall positive effect on facilitating interpersonal mindfulness. However, our analysis contributes insight into the nuanced ways a temporal intervention influences contentious online conversations, pointing to promising areas for future work.

2 RELATED WORK

Social media platforms offer a space for people from diverse backgrounds to engage in political discussions. According to the Pew Research Center, many Americans believe that social media is important for finding others who hold similar views to their own (44.5% of respondents), engaging with others about topics that are important to them (41.5%), and sharing their own political views (38%) [4]. Nonetheless, 55% of U.S. social media users report being "worn out" by online political posts and discussions, according to a 2020 Pew Study of over 10,000 respondents [3]. Studies indicate that arguments over social media are exceedingly common where discussions can become hostile or aggressive, often involving verbal attacks and hurtful behaviour [44, 60]. In fact, *flaming* – the interpretation and sending of aggressive, insulting, or hostile messages in online communication – is common, and may result from de-individuation in computer-mediated communication, characterized by a loss of self-awareness that leads to uninhibited behaviours [33, 35]. Researchers have linked anonymity, lack of eye contact, and perceived invisibility to online de-individuation and disinhibition [1, 33]. Given this, it can be difficult for individuals to engage in civil and genuine conversations about contentious issues on social media, without such conversations becoming psychologically harmful [7, 51].

To provide context on technological approaches to address flaming behaviours in contentious discussions, we situate the present study along four areas of related literature: 1) content-based interventions to mitigate online conflict, 2) design affordances to shape online conversations, 3)

int*ra*-personal mindfulness interventions in social media, and 4) mindfulness interventions in the int*er*-personal domain. We conclude by articulating a gap in the design space.

2.1 Content-based interventions to mitigate online conflict

Currently, the most prominent way that social media platforms address online conflict is through content moderation – that is, restricting what users may post or removing content after it has been posted. Social media platforms routinely employ both human and automated moderation technologies, such as crowd-sourced moderation [34], jury-moderation systems [24], and online communities moderating themselves [64] to remove content that is regarded as offensive, illegal, or otherwise in violation of the platform's terms of service [28]. A 2019 study of the subreddit *ChangeMyView* (reddit.com/r/changemyview/), an online discussion forum promoting productive conversation about polarizing issues, found that content removal is effective for reducing non-compliant behaviour, such as creating posts that violate the rules of an online community [67].

In many cases, however, content moderation has drawbacks. First, content removal is negatively associated with perceived fairness by users, and impedes users' overall experience [57]. This can escalate feelings of frustration among individuals whose content was removed, since users often report lacking a sufficient explanation for content removal, and consequently do not know how to avoid this in the future [79]. Second, automated content moderation can further exacerbate exclusion and inequality online by disproportionately silencing certain groups. In a 2021 paper by Haimson et al. [31], content moderation algorithms were found to disproportionately censor transgender, Black, and conservative social media users. Third, when individuals are banned from popular social media platforms or become frustrated that their content has been removed, research indicates that it may drive them towards more niche platforms with looser moderation policies [82]. Unfortunately, rather than promoting respectful disagreement, such spaces tend to promote one-sided content, often including hate-speech and conspiracy theories [74], which may further reinforce polarization. Finally, when automated content moderation algorithms are inadequate, human content moderation has been found to be damaging to well-being and mental health, as it can expose moderators to dangerous or psychologically disturbing content [68]. Overall, it is clear from the literature that *only* content moderation is not enough to foster civil and respectful online conflict conversation. The next sections foreground an alternative approach: using design affordances within social media to influence and facilitate online contentious conversations.

2.2 Design affordances to shape online conversations

An alternative to content-based interventions is the use of *affordance-based interventions*, where an *affordance* refers to any particular type of interaction that is supported or constrained by the design features of an interface [27, 55]. Social media platforms already make use of design affordances to guide or limit users' communication, such as Twitter's 280 character limit, or WeChat's one-minute voicemails. Such design affordances can directly shape the structure of online conversations and in turn, influence user content. This can have both positive and negative effects. For example, Massanari [48] asserted that Reddit's design features – such as user karma and its sorting algorithm – alongside its policies and culture, facilitated the spread of toxic content, such as revenge porn. Munn [52] found that the ease of rapidly sharing content from Facebook's newsfeed and the algorithmic prioritization of posts that receive high engagement contributes to the proliferation of incendiary, outrage-inducing posts. These examples demonstrate that design affordances within social media platforms can indeed influence online interactions, without necessarily relying on heavy-handed content moderation. There are also positive examples of such influence. For example, in 2017, when Twitter doubled its character limit from 140 to 280 characters, users began to engage in more polite, civil, and constructive online discussions [37]. Micro.blog – a social community of

independent microblogs – omitted sharing and 'like' buttons within their interface to encourage more thoughtful online discussions, rather than simply repeating or amplifying others' posts [59].

Overall, design affordances do not and cannot wholly replace content-moderation; content removal is still employed as an essential moderation tool in these platforms. However, in both mainstream and alternative social media, it seems that design affordances have successfully fostered different types of conversations and user-posted content by guiding behaviours based on what is or is not possible within an interface. Accordingly, users also routinely accept these design affordances when using online communication platforms. This suggests that an affordance-based intervention may be regarded more positively than content-based interventions. The present study explores this line of thought, investigating a design affordance that manipulates the *temporal flow* of a contentious online conversation, specifically by limiting the frequency at which users can send messages to one another while discussing polarizing topics. The goal is to nudge individuals towards more reflective and mindful discussion about controversial issues, rather than engaging in automatic or habitual reactions that may be aggressive, damaging, or unproductive.

2.3 Int*ra*-personal mindfulness interventions in social media

Mindfulness is defined as "paying attention to the present moment, on purpose and nonjudgmentally" [41]. Mindfulness can be studied in an int*ra*-personal or int*er*-personal context. Int*ra*-personal mindfulness refers to mindfulness within the self, while int*er*-personal mindfulness refers to mindfulness during interactions with others. Based on a meta-review by Terzimehic et al. (2019) [73], the literature on mindfulness interventions in CSCW and HCI communities largely focus on the int*ra*-personal domain, with an emphasis on fostering users' *mindful use of technology*. In contrast, our work explores the impact of interventions to foster int*er*-personal mindfulness *between people* during a contentious online discussion. We first present relevant int*ra*-personal interventions to encourage mindful social media use, followed by an introduction to design frictions. We conclude by articulating a gap in the literature: the use of design frictions to foster int*er*-personal mindfulness between strangers during contentious online discussions.

2.3.1 Mindful use of social media. Researchers have explored various interventions for influencing an individual's level of mindfulness when using social media, with the aim of interrupting one's addictive or habitual usage patterns. For example, a study by Song et al. (2021) called "Crank that Feed" required users to crank a physical handle in order to refresh their Twitter feed [66]. Kühn et al. (2019) implemented smartphone features to promote reflective social media use, by reorganizing apps in a single folder to increase difficulty of access [43]. Nudget is a Chrome extension that annotates Facebook feeds in real-time to bring attention to the persuasive or addictive design features at play [75]. These examples highlight a set of technology interventions that aim to encourage more mindful social media use by supporting self-awareness of and reflection on how one engages with technology.

2.3.2 Design frictions. Design frictions are distinct moments of friction or difficulty during an otherwise smooth and efficient user-technology interaction [15, 29]. While designers have traditionally aimed to eliminate "friction" in the user experience due to fear of inconveniencing users, researchers such as Cox et al. (2016) [15] and Gould et al. (2021) [29] argue that intentional moments of interruption in the user experience can, in fact, benefit user well-being. The intention is to shift individuals from habitual, automatic, and mindless user-technology interaction to more mindful and intentional usage. For example, Wang et al. (2014) introduced a delay between when users click "post" on Facebook and when their post is actually published, with the aim of encouraging more intentional posting with regards to user privacy and online self-disclosure [77]. Lyngs et al. (2020) investigated the impacts of a plugin that visually occluded the top portion of the newsfeed

to remind users of their original intention when opening Facebook; their aim was to interrupt mindless scrolling and help users align behaviour with value-oriented goals [46].

Importantly, design frictions differ from interventions to promote mindful social media use, in that they are not built into the *entire* user experience (such as turning a physical crank to refresh a Twitter feed [66]), but are rather time-bounded, distinct "points of difficulty" encountered during users' interaction with a technology [15, p.1390]. The next section will employ the concept of design frictions, shifting from applications in the int*ra*personal to int*er*personal domain.

2.4 Mindfulness interventions in the int*er*-personal domain

Mindfulness in int*er*personal interactions is an essential component of productive conversations [17, 18, 45]. Kabat-Zinn (1993) found that mindfulness promotes attunement, connection, and closeness in interpersonal relationships [40]. Brown and Kasser (2005) correlated mindfulness with a felt sense of social connectedness [10]. Bihari and Mullan (2014) reported that when people related mindfully with others, they experienced a heightened awareness of the tendency to automatically react to triggers (such as distressing interpersonal situations), and developed the mental space to react to others in intentional, rather than habitual ways [8]. Accordingly, people who relate mindfully to others experience increased empathy [17], have better perspective-taking [8], and engage in more effective communication [11].

To date, the literature is sparse with regards to how technology interventions may facilitate interpersonal mindfulness during contentious *online* discussions. One notable exception is the work of Baughan et al. (2021) who proposed *interpersonal design*: a design approach that "centres user relationships in the design process" in order to "support users in the challenging task of arguing well" [7, p. 156]. In Baughan's work, participants engaged in co-design for a series of low-fidelity storyboard sketches depicting various hypothetical design interventions for online arguments. Particularly relevant is Baughan's design concept of *speed bumps*: a mandatory waiting period before users can send (potentially harmful) messages during online conflict. Participant interpretations of this design concept were mixed: some hypothesized that speed bumps may escalate anger, while others reported valuing design features that may encourage them to slow down and be intentional about their behaviour, by "nudging them to pay careful attention to their own words and the words of others" [7].

2.5 The current study

In this work, we build upon Baughan's concept of interpersonal design. We move beyond examining user reactions to storyboard sketches, and conduct a controlled laboratory experiment to investigate the practical implications of a "speed bump" intervention. Specifically, we explore participant experiences with a design friction that limits the frequency at which individuals can send messages to one another during a polarizing online discussion. We distinguish our work from Baughan et al.'s in three ways. First, we explore the actual (rather than anticipated) impacts of this intervention on interpersonal mindfulness through a controlled study. Second, while Baughan's work emphasizes the *outcome* of polarizing discussions (for example, whether a contentious conversation reached a productive conclusion), we instead focus on the *process* of online conflict communication (such as how participants interacted with one another and how they felt during the conflict). Finally, while Baughan's work explores online conflict between individuals who are already acquainted, we focus on conversations between online strangers. We make this distinction because anonymity can increase incivility in computer-mediated communication [30]; therefore, the context of divisive communication between strangers may be most in need of an intervention.

We explore the following overarching research question: *How does limiting the frequency at which individuals can send messages to one another influence their motivation, ability or behaviour to*

communicate mindfully during a contentious online conversation? We anticipate that introducing this temporal design friction may interrupt individuals' automatic and habitual tendencies, giving them an opportunity to reflect on what they write and put more thought into understanding their conversation partner. Consequently, we anticipated that this will foster greater interpersonal mindfulness between strangers during a contentious online conversation. To investigate this, we break down our overarching question into four specific research questions, outlined below.

2.5.1 Specific research questions. We first set out to investigate how limiting the frequency of messages may influence the structure of a contentious online conversation. Previous work highlights how features of a conversation's structure are linked to its content and outcomes. For example, Tan et al.'s analysis of "good-faith" online discussions found that a greater number of words in a message strongly correlates to successful persuasion [71]. Baughan et al. identified that the length of conversations and the time taken to respond correlates to more in-depth (rather than superficial) discussions of contentious topics [7]. Given this, our first research question asks the following:

> **RQ#1**: How do individuals use the mandatory time delay between sending messages, and to what extent does this delay affect the length of their messages and conversation pacing?

Second, we wished to examine the impacts of limiting the frequency of messages on the emotional tone of participants' messages. Key experiences of interpersonal mindfulness include not getting "swept away by a tide of emotional energy" [8] and managing difficult emotional responses when interacting with others [58]. Given this, we pose our second research question:

> **RQ#2**: How does limiting the frequency at which individuals can send messages influence the amount of positive and negative affect in their message content?

Third, we wished to directly investigate the impact of limiting the frequency of messages on int*er*personal mindfulness. Here, we draw upon the *Interpersonal Mindfulness Scale* [58] to understand how participants perceived their own and their partners' levels of interpersonal mindfulness. This scale contains four components (Presence, Awareness of Self and Others, Nonjudgmental Acceptance, and Nonreactivity), though we focus on the latter two, as they are most relevant to computer-mediated rather than in-person communication (detailed in the Method section). We therefore pose our third question:

> **RQ#3**: How does limiting the frequency at which individuals can send messages influence ratings of their own and their partner's level of interpersonal mindfulness (specifically, nonjudgmental acceptance and nonreactivity)?

Finally, we wished to broadly examine how participants felt about the intervention itself and their conversation overall. Our fourth research question asks the following:

> **RQ#4**: What were participants' experiences with a design friction that limited the frequency at which they could send messages during a contentious conversation, and how did this influence their perceived satisfaction with the conversation?

3 METHOD

To investigate our research questions, we conducted an exploratory, mixed-methods experiment wherein 40 participants (20 dyads) engaged in a synchronous online negotiation task about polarizing topics. We explore an existing design feature within a popular social media platform that limits the number of messages per specified time interval that individuals can send to one another, and explore the impacts of this feature during a contentious discussion. Using a between-subjects design, participants were randomized into one of two conditions: the Control Group received no intervention, while the Experiment Group received the design friction intervention. To compensate participants in appreciation for their time, each participant was entered into a randomized draw

to win one of ten $50 gift certificates. This study was approved by the [Anonymized University Name] Research Ethics Board.
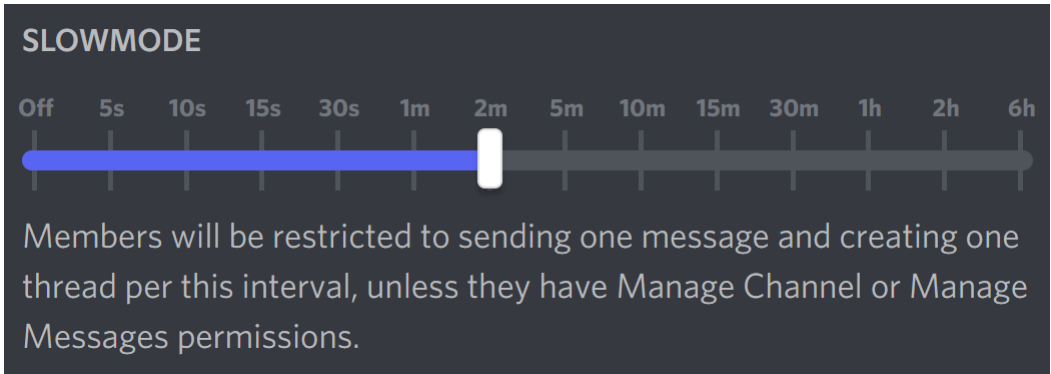
3.1 Discord's "SlowMode" feature

Fig. 1. Discord's SlowMode feature. This study set the slider to one message per two minute interval.

The study took place on Discord, a popular social media platform with more than 150 million monthly active users [21]. Discord allows users to create public or private (invite-only) *servers* for others to join, where members can interact in *channels* through text-based chat or voice. Discord users use this platform to debate and discuss contentious issues, with nearly three thousand public servers labelled with the tag *Debate(s)* [22].

Of particular interest to our study is Discord's *SlowMode* feature, which allows server administrators to limit the number of messages members can send per time interval (see Figure 1). According to the Discord website, SlowMode was intended to "make your channel chill out" [20], for use when multiple users are simultaneously posting to a channel. In this work, we investigate SlowMode in a new context – how this design friction may influence interpersonal mindfulness when dyads (rather than multiple users) discuss polarizing topics within a private server. We chose to focus on private, two-person conversations to isolate the effects of the SlowMode intervention, and because previous literature indicates that social media users value the ability to debate contentious issues in private, rather than on public channels that may have an audience [7]. To ensure that SlowMode would introduce a meaningful delay for the Experiment Group, we chose an interval of one message per two minutes. This was calculated by taking the average time between messages (87 seconds) from the first five dyads of the Control Group, and then setting the SlowMode interval slightly longer than the average time (2 minutes).

3.2 Negotiation task

3.2.1 Task type. To investigate the influence of SlowMode, we have simulated a situation that often arises on social media platforms where two users synchronously argue about polarizing topics on which they disagree [7, 71]. Such situations can be seen in comment sections when users respond to each other in real-time, or on social media platforms that facilitate instant messaging, such as Discord or WhatsApp.

The type of task performed during the study can significantly affect group dynamics and processes [70]. Therefore, to increase the likelihood that participant dyads will encounter disagreement, we chose a "cognitive conflict" negotiation task based on McGrath's Task Circumplex Model [49]. Like

real-world arguments on social media, this task has no objectively correct answer. Rather, partners must work interdependently to resolve conflicts in personal values, attitudes, or beliefs [49].

Specifically, dyads were tasked with an adaptation of the Legislative Dilemma Task [50]. They were presented with three competing advocacy initiatives and asked to come to an agreement on how to allocate $1.8 million ($1,800,000) of hypothetical funds between them. Dyads were told that these hypothetical funds would be used throughout the following year to raise awareness in support of the chosen initiatives. To minimize discussion around how much funding a year-long advocacy campaign would realistically need – as exhibited during pilot studies – and instead focus the conversation on which advocacy initiatives deserved funding, task instructions explained that at least $0.5 million would be needed for funding to be effective. Dyads were additionally told that they each partner has "the same decision-making power" in this hypothetical situation, and were asked to "stay true to [their] values and beliefs" throughout the conversation. To control for differences in participants' sociopolitical environment and the salience of the polarizing topics discussed in the task, we limited recruitment to individuals who reside in Canada, are fluent in English, and are over 18 years old.

3.2.2   Polarizing topics in the negotiation task. The advocacy initiatives that dyads were asked to discuss were chosen based on pilot testing and respondents' answers in the recruitment questionnaire. All topics were highly salient issues in Canada at the time of recruitment and data collection (August to November, 2021): the COVID-19 pandemic was prevalent throughout Canada, with some provinces implementing a controversial "vaccine passport" requirement for people to enter stores and businesses; a growing national opioid epidemic prompted contentious debate around possible policy solutions; and economic concern around the implementation of Universal Basic Income continues to be highly contested among politicians and citizens. Based on this, the negotiation task centred around the following three initiatives:

- Schools, businesses, and other institutions throughout Canada should have the right to ban anybody who has not received the COVID-19 vaccine. *[COVID topic]*
- Using any narcotics should be decriminalized throughout Canada. *[Narcotics topic]*
- Universal Basic Income should be implemented throughout Canada. *[UBI topic]*

3.3   Participants

3.3.1   Recruitment and screening. Participants were recruited through online communities dedicated to discussing controversial political and societal issues, such as relevant Discord servers, Reddit communities, and Facebook groups. Examples include the subreddit, r/CanadianConservatives, and the Discord server, *Increments*, where members discuss polarizing political issues such as COVID-19, climate change, or animal welfare. Recruitment also took place through snowball sampling and social media advertisements.

The study was advertised to participants as a study about "how social media design can better facilitate conversations about polarizing topics between strangers with diverse perspectives." Interested individuals were asked to complete an online recruitment questionnaire, where respondents indicated on an 11-point Likert scale how strongly they agreed or disagreed with four polarizing issues in Canada: COVID-19 vaccinations, narcotics use, universal basic income, and climate change policies. A fifth topic was originally included (*"Refusing to use somebody's preferred pronouns should be considered hate speech in Canada"*) but was removed after an ethics complaint was filed by a respondent during the recruitment phase due to the sensitive nature of this topic. [1]

---

[1]6 dyads were run in the Control Group with the Pronoun topic, COVID-19 topic, and UBI topic. After removal, the pronoun topic was replaced by the Narcotics topic in the negotiation task.

Table 1. Participant demographics

Age	N (%)	Gender	N (%)
18-24	1 (2.5%)	Female	26 (65.0%)
25-34	10 (25.0%)	Male	13 (32.5%)
35-44	18 (45.0%)	Prefer not to say	1 (2.5%)
45-54	6 (15.0%)		
55-64	4 (10.0%)		
65+	1 (2.5%)		

We screened the recruitment questionnaire responses by removing bot-responses using criteria defined by Storozuk [69], as well as respondents whose IP addresses were outside Canada. We then excluded respondents whose views were too moderate to identify points of conflict (that is, individuals who rated one or zero topics with a Likert score greater than 6, or less than 4 on a scale of 11). The *COVID*, *Narcotics*, and *UBI* topics yielded the greatest variability across participants' Likert scores (SD = 3.53, SD = 3.41, and SD = 3.41, respectively, with the Climate Change topic at SD = 3.24). Therefore, these three topics were used to pair the participants to increase the likelihood of conflicts in values during the negotiation task. After filtering for bot-responses, responses from outside Canada, and those who held moderate views, 139 responses remained. Selecting a final sample relied on finding pairs of participants who disagreed on at least two topics, and had overlapping availability to participate synchronously.

3.3.2 Participant demographics. Our final participant pool included 40 participants, randomized into 20 dyads. 32.5% identified as male (n=13), 65% female (n=26), and 1 participant preferred not to disclose their gender. The mean age was 41.3 years (SD=10.6). The Control Group included 16 females and 4 males, while the Experiment Group yielded 10 females, 9 males, and 1 participant with undisclosed gender.

3.3.3 Dyad composition. To ensure dyads would encounter disagreement during the negotiation task, we paired participants who disagreed on at least *two* of the three topics in the negotiation task. We identified *disagreement* based partners' recruitment questionnaire responses, when one partner indicated a Likert score greater than 6 for a particular topic, and the other indicated a Likert score less than 4 for that same topic, on a scale from 0 to 10.

3.4 Experiment procedure

3.4.1 Task environment. The negotiation task was synchronous and text-based, where dyads had 30 minutes to come to an agreement. This time frame was chosen based on pilot feedback, where participants expressed that 30 minutes was the maximum time they would engage in a synchronous online task with a stranger before losing interest, while retaining a sense of response urgency. Dyads in both conditions received the same task instructions, with the exception of an additional sentence in the Experiment Group indicating that "SlowMode is enabled" and that "Each person will be limited to sending 1 message every 2 minutes in this channel." Since our study focuses on the *process* of communication rather than the outcome of the negotiation, to ensure that the 30-minute time limit would not impede dyads' abilities to deeply deliberate on topics they found most engaging, instructions also stated that it was okay if participants could not reach a mutual agreement. To ensure full participant privacy and anonymity, participants used researcher-created Discord accounts that were recycled throughout the study (P1-P2: Control Group, P3-P4: Experiment Group), instead of their existing Discord accounts and usernames. At the beginning of each Discord

conversation, a researcher with the username "[University]Researchers" sent a message indicating that the task has begun. After each participant had sent their first message, the researcher changed their status to "invisible," thus appearing to be offline to minimize participants' feelings of being observed. After 30 minutes had elapsed, or when participants typed "@[University]Researchers, conversation finished," the researcher would appear online again to thank the participants and conclude the task.

3.4.2 Post-task questionnaire. Upon finishing the negotiation task in Discord, participants were asked to complete a post-task questionnaire that investigated participants' level of satisfaction with the conversation, as well as their perceived levels of interpersonal mindfulness. We adapted Pratscher et al.'s Interpersonal Mindfulness Scale (IMS) [58] to form a 20-item survey with questions from two dimensions of the IMS: *Nonjudgmental Acceptance* (defined by Pratscher et al. as "listening without judgment and accepting interpersonal experiences as they occur"), and *Nonreactivity* (defined as "taking time to respond instead of thoughtlessly reacting to another person"). Questions were duplicated and slightly reworded to assess how participants perceived both themselves and their partner. Examples of perceived *self* Nonjudgmental Acceptance questions include "During the discussion, it was difficult to accept that my partner had opinions that were different from mine." Examples of perceived *partner* Nonjudgmental Acceptance include "During the discussion, I found that my partner had difficulties accepting my opinions that were different from theirs." Examples of perceived *self* Nonreactivity questions included "I took the time to form my thoughts before typing out my messages." Examples of perceived *partner* Nonreactivity included "I felt that my partner took the time to form their thoughts before typing out their messages." Participants indicated how frequently or infrequently they encountered such experiences via five-point Likert responses. Any response could be further supplemented with an open-ended comment (for example, if participants wished to refer to concrete quotes from their Discord conversation).

3.4.3 Semi-structured interviews. After participants completed the post-task questionnaire, we randomly selected 20 participants for one-on-one, semi-structured virtual interviews. Of these participants, 13 were available. Of the Control Group interviewees, 5 were female and one was male. Of the Experiment Group interviewees, 4 were female, 2 were male, and 1 preferred not to disclose. All interviews took place over Zoom, and were audio- and video-recorded with a live transcription. Interviews lasted from 23 to 66 minutes (M = 34 minutes).

To reduce the chance of interviewer bias, interviews were conducted by two independent interviewers (the first two authors of this paper). Both interviewers followed the same protocol. Questions common to both study conditions included participants' motivations for joining the study, expectations of the study experience prior to beginning the task, impressions of what went well or not well during the conversation, and comparisons with personal experiences on social media. Experiment Group interviews explored additional questions regarding participants' impressions of SlowMode and how they felt it did or did not influence the conversation.

3.5 Data analysis

3.5.1 Computational text analysis of Discord logs. To investigate participants' written language use in the Discord conversations with regards to interpersonal mindfulness, we employed Linguistic Inquiry and Word Count (LIWC) [56], one of the most highly-cited and validated computational linguistic analysis tools. LIWC relies on a dictionary of 80 categories of content and function words, which reflect diverse psychological processes [72]. Since the literature mapping LIWC language categories to interpersonal mindfulness is limited, our decisions regarding which LIWC variables to measure were based on definitions of *Nonjudgmental Acceptance* and *Nonreactivity* from the Interpersonal Mindfulness Scale [58], along with LIWC category descriptions. As our goal was to

establish a basic measure of the emotional tone of participant messages (rather than the negotiation outcome of the Discord task), we measured LIWC Affect categories, including *overall affect, positive emotion*, and *negative emotion*. For the Affect categories, LIWC assigns a score to a passage of text by counting the percentage of words related to each category. Since LIWC indicates higher reliability for longer (rather than shorter) passages of text [72], LIWC scores were generated by combining each participants' Discord messages into a single passage of text, rather than analyzing individual Discord messages. Additionally, we redacted the term "hate speech" from Discord logs, since this term was present in one of the original topics in the negotiation task (the Pronoun topic). Since participants used this term simply to refer to the topic, LIWC's interpretation that the word "hate" indicated negative emotion was inappropriate.

Additionally, we compared conversational dynamics among each group. We measured the length of participants' messages using LIWC's word count, and measured conversation pacing by calculating the number of seconds between messages for each participant. Differences across all of these measures were investigated using Mann-Whitney U tests.

3.5.2 Post-task questionnaire. Individual responses related to interpersonal mindfulness were combined as per Pratscher et al.'s guidelines for using the Interpersonal Mindfulness Scale [58]. We then used Mann-Whitney U tests to evaluate if there were differences in how members of the Experiment and Control groups self-reported their interpersonal mindfulness and their perceived satisfaction with the conversation.

3.5.3 Demographic analysis. We investigated whether gender or age were significantly related to any of the dependent variables from the computational analysis and post-task questionnaire. To do so, we used non-parametric kernel regression to conduct multivariate analyses, including gender and age as covariates. Age was slightly negatively associated with word count per message and slightly positively associated with message affect, although these associations did not change the results about relationships between the SlowMode intervention and these dependent variables. No other associations were identified. Full results of these multivariate regressions are included in this paper's supplementary documents.

3.5.4 Qualitative analysis. All interviews were fully transcribed by an automatic transcription tool. We then employed an inductive, initial-coding method [61] to identify themes in the Discord conversation logs and participant interview recordings. Four independent coders (the authors of this paper) first individually coded the Discord logs and interview recordings, and identified initial tags for thematic categories. Next, using the process of affinity diagramming [32, 63], team members met weekly using a shared virtual whiteboard to refine, re-organize, and integrate tags to identify commonalities and differences between themes.

The initial codes were organized into the following categories: *Expectations about the conversation, Self-preparation for the conversation, Actual impressions of the conversation, Experiences with current state of social media, Self-disclosure, Pleasantries/friendliness, Subjective and objective language patterns,* and *Agreeing and empathizing.* Through multiple rounds of collaborative discussion over a period of five months, the authors iteratively revised and agreed upon a final taxonomy of codes organized into thematic categories. The final categories were: *Rapport-building, Stop-and-think, Message impact, Keeping on track, Frustration, Productivity,* and *Combativeness.*

3.6 Limitations

With this study design, we aimed to understand how a temporal design intervention (SlowMode) can influence participants' subjective experiences during a contentious online conversation. The

study design was controlled, focusing on dyadic interactions in a private communication context. Accordingly, our limitations concern the generalizability of this study task to polarizing conversations as they might occur in typical social media encounters.

First, sampling bias may have limited generalizability. Although participants did not know the task instructions before joining the study, they likely anticipated engaging in a polarizing conversation with another participant, due to how the study was advertised. Yet, they still decided to join the study and commit to engaging in such a conversation. This suggests that study participants may have been less conflict-averse than the general population. Additionally, we filtered out participants who held views that were too moderate to identify points of disagreement with others. Therefore, study participants may have also been more strongly opinionated than the general population. In typical social media encounters, however, users can have relatively moderate views, and may not always expect conflict nor are they necessarily prepared for or committed to engaging in a full discussion with a stranger for a specified length of time. While strong opinions in this sample likely magnified the potential for conflict, participants' willingness to commit to the study task may have led to more mindful conversations overall. Importantly, these characteristics would be present in both the Control and Experiment Groups, and therefore would not have compromised internal validity.

Second, dyadic interactions took place in a private chatroom, with only the researchers observing the conversation rather than a public audience, as in typical social media. The intention of the private communication environment was to better control for and isolate the effects of the SlowMode intervention. Yet, the private setting may have also influenced interpersonal mindfulness. For example, without an audience watching, participants may have been less concerned with seeking approval from onlookers or engaging in performative aspects of online conflict, which may have increased interpersonal mindfulness. Alternatively, without an audience, participants may have felt less inhibited when responding and less concerned about managing a public persona, which may have decreased interpersonal mindfulness.

Finally, participants were completely anonymous to one another during the study. This differs from some social media experiences, where users often publish a profile picture or other identifying information. Prior research has identified that online anonymity can both increase the potential for *flaming* by decreasing a sense of humanity between partners, as well as reduce the potential for prejudice or stereotyping based on race, gender, or other visible features [35]. Future work, focusing on platforms on which users are identifiable, may consider modifying the research design accordingly.

Overall, the features of our study design likely set this experience apart from participants' typical experiences on social media, possibly influencing interpersonal mindfulness and conversational dynamics. However, these effects applied to both the Experiment and Control study conditions. Therefore, internal validity was upheld, and comparing the two conditions cancels out the possible effects of participants' strong opinions and task commitment, and the study environment's privacy and anonymity, allowing us to pointedly investigate the effects of the SlowMode intervention. Although we may not be able to draw conclusions about the greater social media landscape, the present study offers a starting point to examine how a temporal design intervention operates within a dyadic, online contentious conversation.

4 FINDINGS

To address our four research questions, we triangulated quantitative and qualitative data from analysis of the Discord logs, recruitment surveys, post-task questionnaires, and participant interviews. The following sections discuss the influence of SlowMode on various components: conversational dynamics (*RQ#1*), emotional affect (*RQ#2*), interpersonal mindfulness (*RQ#3*), and participants'

perceptions of the intervention (*RQ#4*). For each, we present results from quantitative analysis, followed by results from qualitative analysis, which offer a deeper contextual understanding of the intervention's influence and help to explain the quantitative findings.

Throughout, we refer to participants by their dyad number, participant number, and study condition, where "SM" stands for SlowMode. For example, "D6P1" refers to Dyad 6, Participant 1 in the Control Group, while "D13P4-SM" refers to Dyad 13, Participant 4 in the Experiment Group. P1 and P2 names were recycled throughout the study for the Control Group, while P3 and P4 were recycled for the Experiment Group. To protect participant anonymity, we employ the use of singular "they" pronouns when referring to participants. This approach aligns with recent gender-neutrality discussions in the Human-Computer Interaction community [12], and recent APA guidelines [5].

4.1 Conversational dynamics

This section addresses our first research question:

> **RQ#1**: *How do individuals use the mandatory time delay between sending messages, and to what extent does this delay affect the length of their messages and conversation pacing?*

By analyzing the Discord conversation logs, we found that introducing the two-minute SlowMode delay had significant effects on message timing, total amount of text sent, and individual message length. Supplementing these quantitative findings with qualitative interview analyses, we identify that these changes in conversational structure resulted from participants making use of the two-minute delay to more carefully construct their messages before sending.

4.1.1 Log analysis - conversation pacing. Figure 2 illustrates a comparison of the mean intervals between participant messages in the Experiment and Control Groups, as analyzed from the Discord conversation logs. As expected, participants in the Experiment Group (Median = 258.44 seconds, Mean = 258.43, SD = 76.16) took significantly longer between messages, compared to participants in the Control Group (Median = 87.09 seconds, Mean = 87.82, SD = 32.19). A Mann-Whitney U test confirmed that the difference was significant (z = -5.36, U = 2, p = 0.000) and that there was only a 1% likelihood that a participant in the Experiment Group would have taken longer between messages than a participant in the Control Group. Notably, in the Experiment Group, the difference in timing exceeded the mandatory two-minute SlowMode delay, where, in fact, participants took on average over four minutes between sending messages.
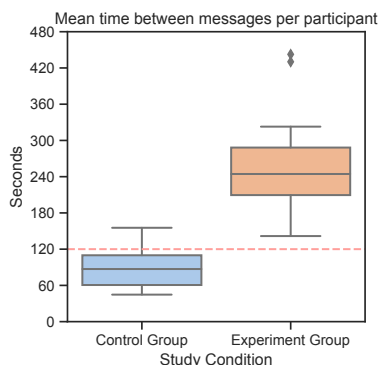

Fig. 2. On average, participants in the Experiment Group had a mean of 258 seconds between messages, compared to 88 seconds in the Control Group. The dashed red line indicates a 120 second SlowMode threshold.

4.1.2 Log analysis - message length. Figure 3 illustrates a comparison between the amount of text written by participants in each study condition. Overall, participants in the Experiment Group wrote significantly less text (Median = 346.0 words, Mean = 324.35, SD = 89.01) than individuals in the Control Group (Median = 489.0 words, Mean = 459.45, SD = 206.19). Mann-Whitney U test results indicated a 71% likelihood that a participant in the control group wrote more overall text than one in the experiment group (z = 2.25, U = 117, p = 0.025). However, this finding was reversed when it came to individual message length. On average, Experiment Group participants wrote longer individual messages (Median = 33 words, Mean = 41.56, SD = 32.79), compared to the Control Group (Median = 13.5 words, Mean = 19.89, SD = 19.83). Mann-Whitney U test results indicated a 29% likelihood that a given message written by a control group participant was longer than one written by someone in the experiment group (z = -7.83, U = 19640.5, p = 0.000).

Lastly, despite the fact that Experiment Group participants exchanged less text overall, the task-completion rate was the same across both the Control and Experiment Group. In both groups, 6/10 dyads successfully completed the task by reaching an agreement.
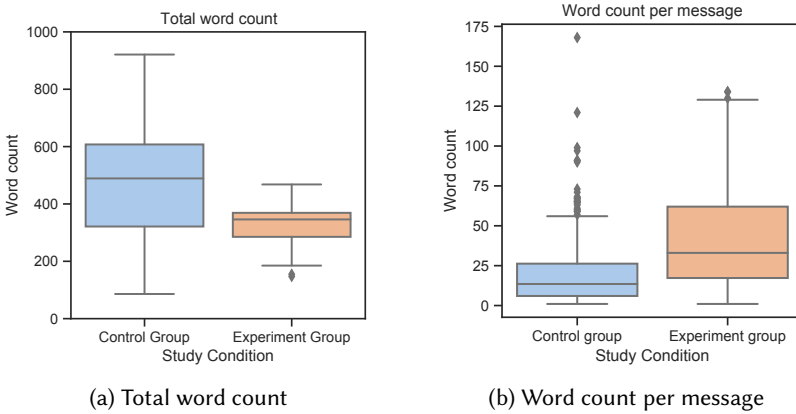

(a) Total word count   (b) Word count per message

Fig. 3. Participants in the Experiment Group wrote fewer words overall, but significantly more words per message.

4.1.3 Interview analysis. Qualitative interview analysis supplements these quantitative findings by revealing how Experiment Group participants used the SlowMode delay, and why their messages were longer than in the Control Group. 6/7 interviewees in the Experiment Group reported using the time afforded by the SlowMode delay to more carefully construct their messages before sending. Some participants used this time to ensure that each message was impactful and meaningful, as described in the following example:

> D7P3-SM: *"I really had to put a lot of thought into that one comment [...] because, once I hit that enter button, it was two minutes before I could say anything else. To keep the conversation going, you want to say meaningful, insightful, productive things."*

For others, SlowMode seemed to encourage more critical evaluation of their messages, though it may not have influenced actual message content in significant ways.

5/7 Experiment Group interviewees explicitly noted that longer message lengths and intervals between messages directly resulted from a more careful construction of their messages before sending. For example, D9P4-SM stated that they wrote *"larger ideas rather than short little yes or*

Table 2. Mann-Whitney U tests showed no significant differences in LIWC's affect scores between the experiment and control groups

	Control group			Experiment group			Mann-Whitney results		
	Median	Mean	SD	Median	Mean	SD	z	U	p
Affect	6.29	6.40	1.29	5.89	5.65	1.31	1.65	139	0.099
Positive affect	5.18	5.06	1.04	4.36	4.58	1.28	1.68	138	0.090
Negative affect	1.13	1.25	0.63	0.93	0.98	0.90	1.58	142	0.113

no responses," while D7P3-SM noted that *"As soon as you hit Enter, you had to wait to say anything else. So, if you were going to be saying one-word answers, you weren't going to be saying a whole lot."* Finally, other interviewees in the Experiment Group indicated that by ensuring each message was more *"well-written and well-understood"* [D3P4-SM], their writing time inevitably extended beyond the two-minute SlowMode minimum. In contrast, Control Group interviewees did not comment on actively taking time to put a degree of care into constructing thoughtful messages.

Overall, the SlowMode intervention significantly changed the pace and structure of participants' conversations in the Experiment Group, increasing the depth and purpose with which participants approached writing their messages. This offers an explanation for why the task-completion rate was the same in both groups, even though Experiment Group participants were constrained by SlowMode and exchanged less text throughout their conversations.

4.2 Emotional affect

Our second research question asked the following:

> **RQ#2**: *How does limiting the frequency at which individuals can send messages influence the amount of positive and negative affect in their message content?*

4.2.1 LIWC analysis. To answer this quantitatively, we relied on LIWC computational text analyses of the Discord logs for the category "affect." Table 2 reports summary statistics and Mann-Whitney U test results comparing LIWC affect scores between the Control Group and the Experiment Group. The summary statistics indicate that participants in the Experiment Group used slightly fewer affect words than those in the Control Group. However, a Mann-Whitney U test did not find this difference to be statistically significant (p = .099).

4.2.2 Interview analysis. Qualitative interview analysis provides a more nuanced answer to this research question when comparing the experience of Control and Experiment Group participants. When asked about their general impressions of their conversations, all Control Group interviewees described perceiving an overall "friendly" tone during their conversation. This may have resulted from their choosing to exchange pleasantries and build rapport at the beginning of the negotiation task – a behaviour that occurred spontaneously and was not included in the task instructions. D8P2 described this experience:

> [D8P2]: *"[My partner] already set the tone as a friendly, inviting person. If the tone was different, maybe the conversation would've gone differently, and we would've gotten defensive. But we didn't. We both had our opinion and we respected each other's opinion."*

On the other hand, when asked about general impressions of their conversations, Experiment Group interviewees focused on discussing the efficiency of their communication rather than pointing out any particular emotional tones. All interviewees in the Experiment group commented that SlowMode introduced a time pressure during the conversation, which encouraged them to

stay focused on the conversation goal and exchange fewer pleasantries at the beginning of their conversations, as described in the following example:

> [D9P3-SM]: *"At the beginning, we were doing the nice, 'Hey, how's it going?' and then we were like, 'Oh, this isn't going to get us anywhere.' I knew we had a finite amount of time to get the task done, and that wouldn't have been very productive."*

Similarly, D7P3-SM commented on the potential for this intervention to help people focus on the goal of the conversation, rather than *"getting carried away, [...] going back and forth, and then you're kind of getting lost in your own conversation."* These interview quotes suggest that while Control Group interviewees were more likely to focus on emotional bonding when giving their impressions of their conversations, Experiment Group interviewees focused on discussing the task itself. Thus, the role of affectual messages may have been diminished in the Experiment Group.

Some evidence also suggests that when a conversation is nearing hostility, SlowMode may help to mitigate the risk of emotional escalation. 5/7 interviewees in the Experiment Group hypothesized that SlowMode may lead to less emotionally charged discussions in settings with greater hostility than they encountered in this study, as exemplified below:

> [D9P4-SM]: *"There were definitely times when I wanted to say something that I couldn't say. So I actually think it would be positive, in that if we were getting heated, [SlowMode] would have slowed us down and made us think about what we were typing ahead of time."*

One interviewee in the Experiment Group (D3P4-SM) believed their conversation did, in fact, have the immediate potential to become hostile, particularly when their partner accused them of supporting segregation of Canadians due to their support of mandatory COVID-19 vaccinations. Here, D3P4-SM used the SlowMode delay to "calm" themselves and refine their drafted message, while trying to maintain a productive conversation:

> [D3P4-SM]: *"It was shocking [...], like, 'What the hell are you talking about?' I had to spend a minute to calm myself and think, 'How can I communicate with this individual? How can we complete this task knowing that [...] this is the assumption they made about me?' That's why it took so long. [...] There were a couple times where I had to go back and delete what I was typing, because I realized this is more me having an argument, rather than trying to come to some sort of conclusion or understanding."*

In summary, our results suggest that SlowMode introduced a time pressure that may have inhibited emotional affect during conversations, by urging participants to send more efficient, task-focused messages. Experiment Group interviewees anticipated that SlowMode may have a hostility-reducing effect in hypothetical heated conversations, while one participant actually perceived hostility in their Discord conversation, and used the SlowMode delay to help deescalate. In contrast, Control Group interviewees perceived a positive tone throughout their conversations, possibly due to their ability to freely build rapport and exchange pleasantries without being (as) cognizant of time.

4.3 Interpersonal Mindfulness

Our third research question was as follows:

> **RQ#3**: *How does limiting the frequency at which individuals can send messages influence ratings of their own and their partner's level of interpersonal mindfulness (specifically, nonjudgmental acceptance and nonreactivity)?*

4.3.1 Questionnaire analysis. Although participants' median rating of their own nonreactivity was slightly higher in the Control Group than the Experiment Group, analysis revealed no statistically

significant differences in this or other dimensions of the self-reported interpersonal mindfulness ratings between the Experiment and Control Group (See Table 3).

4.3.2 Interview analysis. Analysis of our interview responses provides context for these quantitative data. Interviewees in both the Control and Experiment Groups indicated that characteristics of the study setup may have influenced their self-perceived and other-perceived interpersonal mindfulness. Participants mentioned the likelihood of this study to recruit people who were already motivated to engage in a respectful polarizing conversation with a stranger. For example, D8P1 said, *"It was a very intentional process, the people who would've participated would have had an expectation of themselves and who they were speaking to, to be respectful."* Similarly, D9P3-SM expected this study to involve a *"respectful conversation where someone will debate with you."* Together, these participant expectations of the study may have led to a ceiling effect of interpersonal mindfulness across both groups, possibly explaining the similarity in quantitative measures of nonjudgmental acceptance and nonreactivity.

Additionally, 3/7 Experiment Group interviewees indicated that aspects of their personality may have caused them to be resistant to the effects of SlowMode. D3P3-SM, who described themself as primarily using "logic-based reasoning" during contentious conversations, remarked that *"if I was prone to [writing] emotional responses, [SlowMode] would give me time to reflect on it."* D9P3-SM, who described themself to be "already very careful" when discussing contentious topics, stated, *"because I am so conscious of what I say in controversial things, I don't know if [SlowMode] would have made a huge difference."* Finally, D21P4-SM indicated that they typically take *"a long time to respond"* and that *"[SlowMode] almost slowed me down more, because I was worried about sending something quickly."* Overall, these comments indicate that personality traits of partners may have further mediated the influence of the SlowMode intervention.

Although the study set up, and participants' personalities, may have influenced quantitative measures of interpersonal mindfulness, thematic analysis of interview data revealed a distinct "stop-and-think" theme in the Experiment Group that was not present in the Control Group. 5/7 Experiment Group interviewees commented on the particular influence of SlowMode to urge them towards more thoughtful and productive communication. For example, D21P3-SM stated, *"[SlowMode] really forced me to stop and think, 'Okay, what's coming my way, and what am I going to say [...], instead of just blurting out."* D9P4-SM said, *"[SlowMode] made us think about our feedback and what we were putting in."* Quotes like these from the Experiment Group suggest

Table 3. Mann-Whitney U tests showed no significant differences between interpersonal mindfuless scales between the experiment and control groups

	Control group			Experiment group			Mann-Whitney results		
	Median	Mean	SD	Median	Mean	SD	z	U	p
Ratings about self:									
-Nonjudgemental acceptance	15.00	13.95	2.52	14.00	13.55	2.35	0.61	178	0.550
-Nonreactivity	21.00	20.15	3.27	18.00	18.10	4.01	1.63	140	0.100
Rating about partner:									
-Nonjudgemental acceptance	15.00	14.45	2.01	14.50	13.60	2.98	0.72	174	0.471
-Nonreactivity	20.50	20.20	4.05	21.00	19.35	4.91	0.44	184	0.664

that the SlowMode intervention facilitated nonreactivity (*"it forced me to stop and think"*), and nonjudgmental acceptance, as participants prepared for *"what's coming their way."* Control Group interviewees made no comparable comments about actively "stopping and thinking" during their conversations.

4.4 Experiences and perceptions of the SlowMode intervention

Finally, our fourth research question asked the following:

> **RQ#4:** *What were participants' experiences with a design friction that limited the frequency at which they could send messages during a contentious conversation, and how did this influence their perceived satisfaction with the conversation?*

4.4.1 Questionnaire analysis. In the post-task questionnaire, participants responded to the question "In general, how satisfied do you feel with your conversation over Discord?" on an 11-point Likert scale. Figure 4 suggests that Experiment Group participants reported lower satisfaction (Median = 8.02, Mean = 7.143, SD = 2.558) than individuals in the Control Group (Median = 8.90, Mean = 8.587, SD = 1.433). However, a Mann-Whitney U test showed that members of the Experiment Group were not significantly more likely to have a low satisfaction rating than members of the Control Group (p = .055). Since this result is borderline, and Figure 4 shows that the minimum satisfaction scores were substantially lower in the Experiment Group, it is possible that a larger sample may have identified a statistically significant relationship. This is an important possibility to consider because an intervention that frustrates users could potentially backfire and contribute to elevated hostility [7].
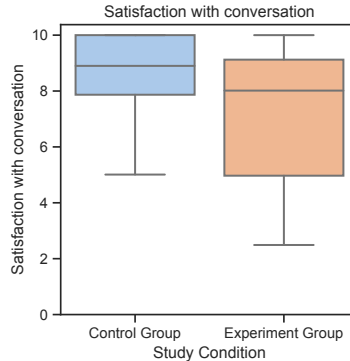

Fig. 4. Participants in the Experiment Group reported lower satisfaction with the conversation than participants in the Control Group.

4.4.2 Interview and log analysis. From the interview data and conversation log analyses, we identified three factors that may have informed participants' satisfaction and their overall experience with the intervention. SlowMode seemed to: 1) frustrate participants, likely lowering satisfaction; 2) keep them focused on the task, which they viewed favourably; and 3) change the method by which dyads built rapport with one another.

First, all Experiment Group interviewees described the SlowMode intervention to be either "frustrating" or "annoying." Frustration stemmed from one or more of the following factors: being required to wait before sending a message, having to include many future arguments in one message, and not being able to correct a mistake or misunderstanding until two minutes had passed. While

most participant criticisms of SlowMode related to its disruption of the natural conversational flow, D3P3-SM, who opposed any form of *"social media trying to regulate what is said,"* stated that SlowMode was *"big-brothery"* and *"offensive in that we were being limited."*

Second, 6/7 Experiment Group interviewees appreciated that the SlowMode intervention encouraged them to focus on the goal of their conversation and communicate efficiently without getting sidetracked by other topics. For example, when describing the intervention, D7P3-SM explained, *"It was good in the way that it kept things on track, it didn't allow us to jump around too much or get too carried away with things."* The Discord logs also support this notion. For example, in D3-SM, the third message in the Discord conversation was, *"So we are limited on time so let's get right down to it yea? Out of these 3, how would you place order of importance?"* Similarly, in D9-SM, the fifth message was, *":) This one message every two minutes thing will be tricky. Okay. So we have a job to do! What are your preliminary thoughts?"*

Further analysis of Discord logs support that Experiment Group dyads tended to "get right down to it" when compared to the Control Group. Notably, Control Group dyads sent an average of 5.20 pleasantry messages before the first task-related message, compared to 2.70 pleasantry messages in the Experiment Group. A Mann-Whitney U test showed that there was a 90% likelihood that a given dyad in the Control Group would exchange more pre-task pleasantry messages than a dyad in the Experiment Group ($z = 3.413$, $U = 10$, $p = 0.001$). Additionally, after initial pleasantries were sent in Experiment Group dyads, 100% of their messages were task-focused and made progress towards an agreement. That is, although some messages contained rapport-building sentences, zero messages were entirely unrelated to the task. In the Control Group, however, 8% of the messages sent during the main negotiation were entirely unrelated to the task. This may explain why Experiment Group interviewees described themselves as more focused. Since they wrote longer messages and, after minimal initial pleasantries, all messages were focused on the goal of the conversation, they did not have quick throwaway messages that could have taken them off-track.

Task-focus was also reflected in how readily participants shared their opinions about *all* of the polarizing topics. At least one partner in 9/10 Experiment Group dyads stated their beliefs about all three topics within their first task-related message. In contrast, only 4/10 Control Group dyads did this, and the other six dyads stated only one opinion in their first task-related message, and revealed beliefs about the other topics gradually.

Finally, SlowMode changed the way partners built rapport with one another. In the Control Group, participants built rapport through exchanging pleasantries at the beginning of their conversations, and social messages including jokes, such as *"I wish I had that kind of money to give away!"* [D10P1], and voluntary admissions, such as *"Btw, math is NOT my strong suit"* [D10P1] and *"I'm not a fast typer"* [D6P1]. In contrast, for Experiment Group participants, rapport was not built through the explicit exchange of pleasantries and off-topic comments, but rather the shared experience of learning how to work with and collaborate within the constraints of slow mode. For example, D15P4-SM wrote, *"Oh this slow mode is going to be a killer haha,"* where their partner responded *"Yeah I'm also a little confused about the slow mode haha."* Similarly, D21P3-SM first wrote, *"Ok this slow mode will be slooooow,"* followed by, *"Ouch! I pushed the sent button too early"*, where their partner responded *"Haha! It's hard to not push send!"* These conversation log quotes indicate that for Experiment Group participants, social bonding was built through the shared experience of navigating SlowMode.

5 DISCUSSION AND FUTURE WORK

Our findings revealed mixed effects of the SlowMode intervention: it simultaneously frustrated participants by disrupting the natural flow of their conversations, while also facilitating more thoughtful and goal-oriented communication. We focus our discussion around two overarching

dichotomies. First, Experiment Group participants expressed frustration towards the two-minute SlowMode delay, while also appreciating a heightened perceived importance of each message, leading to more reflective message construction compared to Control Group participants. These results mirror the findings from Baughan et al.'s (2021) co-design exercises (presented in Section 2.0 Related Work), which explored how hypothetical design interventions could facilitate healthy online arguments [7]. Some of Baughan's participants hypothesized that slowing down an online argument could encourage thoughtfulness and intentionality. Our interview analyses confirm this prediction, indicating that this was a real outcome of such an intervention, when put into practice. Additionally, participants in Baughan's study hypothesized that a mandatory waiting-period could cause preexisting anger to escalate during a conflict. In the present study, participants did experience feelings of frustration. However, these feelings did not seem to escalate conflict towards their partner, and were instead directed at the intervention itself. This combination of frustration and increased thoughtfulness is characteristic of design frictions. Cox et al. (2016) argue that a design friction, although brief and sometimes disruptive, can prompt users to switch from *System 1* behaviours (which are automatic, reactive, and habitual) to *System 2* behaviours (which are intentional actions performed with an active awareness of oneself) [15]. Our findings are consistent with this description, where a short, somewhat disruptive "friction" that limited how often individuals could send messages seemed to have nudged participants from System 1 to System 2 modes of interaction while discussing polarizing topics.

The second dichotomy that arose in our findings is as follows: while the intervention conflicted with Experiment Group participants' desire to engage in social bonding, they appreciated the increased efficiency of communication and greater focus on the goal of the conversation. We unpack this using Social Information Processing Theory (SIPT). Based on SIPT, building trust and relationships in computer-mediated communication (CMC) takes more time than in face-to-face (FtF) communication. [38, 76]. Accordingly, while Control Group dyads seemed able to build rapport with minimal concern about time, SlowMode imposed additional time pressure on Experiment Group dyads that seemingly motivated them to sacrifice rapport-building in favour of task-focus. In addition to taking more time, SIPT states that building relationships in CMC relies on different social cues than in FtF communication [38, 76]. One such social cue in CMC is response time [42]. SlowMode interfered with this social cue, further complicating the ability to build rapport in the Experiment Group. We posit that participants' satisfaction with their conversation may have decreased due to SlowMode's added time-pressure and interference of social cues, and that these effects may offer an explanation for the differences in rapport-building and task-focus across study groups. Meanwhile, this streamlined style of communication was also appreciated by Experiment Group interviewees. To understand such appreciation in spite of SlowMode's disruption of social-bonding, we draw a parallel to Di Blasio and Milani's work on the differences between CMC and FtF communication [19]. Their work indicates that since CMC contains fewer "distractions" than FtF communication, such as ongoing social cues and constant non-verbal stimuli, interlocutors can more easily engage in logical reasoning when debating with one another. We posit that SlowMode similarly limited "distractions" in the Experiment Group compared to the Control Group, by minimizing the social cues that would normally be present in synchronous CMC, possibly leading to a greater sense of ease and ability to negotiate funding allocation and engage with the task.

Lastly, in spite of the patterns described above, our quantitative analyses did not find statistically significant differences between groups in terms of either the level of affect in participant messages (measured using LIWC's sentiment analysis) or in post-task scores on the Interpersonal Mindfulness Scale. One possible explanation for this is, because the effects we observed were mixed, they may not have resulted in a clear increase or decrease in quantitative measures of affect and interpersonal mindfulness. As a whole, we cannot conclude that SlowMode had an overall positive

effect on facilitating interpersonal mindfulness in conflict communication. However, our analysis has contributed initial insights about the tradeoffs between smooth conversational pacing and intentional communication, as well as between social-bonding and goal-orientated communication, which point to opportunities for future work on similar temporal interventions.

Given these overarching dichotomies, we reflect on four directions for future work: i) Embracing design frictions to facilitate challenging conversations, ii) Investigating temporal frictions in conflict-laden communication contexts, iii) Designing for sender- and recipient-oriented focus, and iv) Investigating frictions for ethical persuasion.

5.0.1 Embracing design frictions to facilitate challenging conversations. Conversations about polarizing or contentious issues can be naturally frustrating, awkward, or otherwise difficult [36, 39, 80]. Yet, engaging in such conversations are critical to foster a tolerance for others' diverse political views [53] and to develop complex and nuanced political perspectives [23]. Our findings suggest that embracing design frictions and the frustration they may evoke, may *enhance* rather than impede the meaningfulness of a challenging conversation.

Our findings also suggest that if a topic of discussion is inherently significant or important, interlocutors may persevere through frustration, so as to adequately express themselves or accomplish a meaningful collaborative goal. In such contexts, seamless (frictionless) designs can potentially exacerbate harm by supporting *System 1* communication behaviours that are quick, automatic, or emotionally reactive [15]. For example, previous work has shown that motivations to engage in online flaming stem from the search for immediate gratification [1], which is made possible in part by the *ease* at which contention can arise in online conversations. Furthermore, effortful communication prompted by a novel communication format has been found to elicit greater interpersonal mindfulness [11]. Therefore, we propose that the user experience should not, in fact, be so seamless in contentious conversations. Rather, temporal design frictions such as SlowMode have the potential to remind people of the significance of their discussion, urging users to respect these conversations as meaningful, important, and therefore effort*ful*, rather than fleeting and effort*less*.

5.0.2 Investigating temporal frictions in conflict-laden communication contexts. Temporal design frictions that introduce a minor delay during an otherwise real-time discussion may be especially useful in high-stakes, conflict-laden situations, as they can provide an intermediate solution to facilitating mindfulness in a conversation that is neither fully synchronous nor asynchronous. This study explored the impact of a temporal design friction on interpersonal mindfulness during an online political discussion between strangers. Our findings demonstrate that although the intervention interfered with the ability to build rapport, Experiment Group participants appreciated that SlowMode focused their conversation on the common goal of reaching an agreement. Additionally, negotiations in both groups were generally civil, so this study did not capture explicit flaming, toxicity, or harassment, which are urgent issues in social media. Given these two observations, future work that embraces the use of temporal design frictions may be fruitful in cases where i) interlocutors are already motivated to collaborate towards a common goal with minimal desire to socialize with one another, and ii) the discussion has the imminent potential to become hostile. Such elements can exist in high-stakes contexts such as sensitive legal negotiations or interactions between separated co-parents or partners in hostile relationships.

Indeed, prior work has found that asynchronous communication may be preferred over face-to-face communication by partners in hostile relationships, due to the ability to carefully construct messages, maintain boundaries, document what is being communicated, and limit the necessity to speak directly [26, 47, 65]. Meanwhile, a fully asynchronous solution also has downsides (for

example, in hostile post-divorce contexts) if there is a need for urgent communication (such as child-care planning) [26, 47, 65]. Likewise, CMC may be preferred in attorney-client communication over face-to-face, to overcome time and distance obstacles, communicate with more precision, and give space for a client to digest difficult information [6, 62]. However, the formal nature of asynchronous communication such as email can depersonalize and harm an attorney-client relationship, while fully synchronous interactions can lead to hasty responses, miscommunication, and online flaming [6, 62]. Therefore, temporal design frictions such as SlowMode, which encourage a minor delay during an otherwise real-time conversation, may offer an intermediate solution in these scenarios.

Additionally, future studies could investigate temporal design frictions in online conversations where conflict is artificially controlled for. For example, rather than two participants conversing, future work may benefit from using a "confederate" who covertly acts as one of the interlocutors, following a protocol to introduce conflict (as seen in [16]). This type of study design would allow for more control over the level of conflict, although it would also require rigorous planning to ensure valid representation of real-world online arguments.

5.0.3 Designing for sender and recipient-oriented focus. Notably, none of the Experiment Group participants in this study commented on using the SlowMode delay to more carefully read or interpret *their partners'* messages. Instead, participants primarily discussed how SlowMode influenced their own ability to *construct* and send messages. In contrast, Baughan et al.'s co-design interviews suggested that slowing down an online argument could urge interlocutors to not only write more thoughtful messages, but also more carefully attend to others' messages [7]. Thus, our findings reveal an interesting *sender-oriented* focus (How can I get my message across?), rather than a *recipient-oriented* focus (How can I better understand my partner?). Here, we identify a gap for future exploration. Literature on conflict-resolution [13, 25] and interpersonal mindfulness [58] indicate that empathetic and active listening skills are critical in conflict situations. Thus, future work may benefit from exploring how design frictions can facilitate not only a *sender-oriented* focus, but also a *recipient-oriented* focus to foster deeper understanding and respect. For example, in addition to prioritizing clear or effective communication, the intervention may also target the promotion of empathetic, non-reactive, and non-judgmental interpretations of emotionally-charged messages from others. To achieve this, the present intervention could be modified to introduce a design friction elsewhere in the process of conversing. For example, modifications may include: a delay after *receiving*, rather than after *sending* a message to allow individuals time to read and be mindful of their interpretation; not allowing interlocutors to begin typing until indicating that they have read their partner's message; or revealing messages slowly, perhaps word-by-word, so that mindfulness is encouraged throughout the reading and interpretation process.

5.0.4 Investigating frictions for ethical persuasion. Existing literature suggests that persuasive interventions should align with users' values, to be most effective and perceived as ethical. For example, Baughan et al. concluded that interventions that are too intrusive or "dilute users' intentions" may cause harm during an already hostile exchange. [7], and Branch et al. (2021) found that individuals do not wish to be persuaded by technology "unless they view the motivation of the persuader as morally admirable" [9]. Below we examine the extent to which SlowMode may have aligned with participants' goals, and suggest how to enhance the degree to which future temporal frictions can yield effective and ethical outcomes.

While SlowMode enforced a mandatory delay of two minutes, Experiment Group participants regularly and voluntarily exceeded this delay, taking an average of four minutes to send each message. This extra time suggests that participants did not merely tolerate the intervention, but seemingly embraced the time delay, with the goal of communicating clearly, carefully, and intentionally. Therefore, in contrast to content moderation approaches that correct behaviour after users

have acted or enforce behaviour that may not align with users' values, it seems that this design friction gently *nudged* users towards mindful communication behaviours that they were already motivated to engage in.

Despite participants voluntarily exceeding the two-minute delay, interview analysis revealed that participants were, in fact, frustrated by the delay because it intruded on participants' desires to converse smoothly and engage in rapport-building. Therefore, we suggest two ways to further align a temporal friction with interlocutors' existing values. One way to implement such ethical persuasion and to encourage users to embrace the intervention is to carefully orchestrate the framing of the intervention. In this study, we did not reveal that the purpose of the intervention was to facilitate interpersonal mindfulness. Rather, Experiment Group task instructions stated that "SlowMode is enabled. This means that each person will be limited to sending 1 message every 2 minutes in this channel." This wording frames the intervention as a "limit" or constraint. Future work may benefit from investigating how a positive framing of the intervention, such that it may align with user values (for example, being non-judgmental, tolerant, and accepting), could lead to more positive perceptions of the intervention.

Second, a temporal intervention should be sensitive to the changing state of dynamic conversations. In this study, we investigated the influence of SlowMode when enabled consistently throughout a conversation. While increased thoughtfulness during the main negotiation was well-received and aligned with participants' goals, SlowMode conflicted with participants' desires to exchange pleasantries at the beginning of their conversations. This has two implications. First, future research designs may benefit from more precisely isolating the effects of a temporal intervention around moments of conflict, for example, by allowing an unrestricted period of exchanging pleasantries prior to the main discussion, or by instructing participants to begin the negotiation task immediately, with no opportunity to exchange pleasantries beforehand. Second, in practical applications, adaptive systems may be necessary to accommodate interlocutors' shifting needs and goals throughout a conversation. This may be achieved by incorporating natural language processing, such as sentiment analysis [e.g., 2] and toxicity detection [e.g., 54, 81], which could detect when a temporal intervention may have the greatest positive impact. These avenues for future work can provide insight into how persuasive design frictions should adapt to the goals of users, which is necessary for positive user experience and ethical outcomes.

6 CONCLUSION

We presented the results of an exploratory, mixed-methods experiment investigating the influences of a temporal design intervention on interpersonal mindfulness during a contentious online discussion between strangers. 40 participants (20 dyads) took part in a synchronous negotiation task about three polarizing topics. Dyads were randomized into one of two conditions: the Control Group received no intervention, while the Experiment Group was limited to sending one message per two minute interval. Findings revealed a mixed effect of the intervention, simultaneously frustrating participants while also facilitating more intentional, thoughtful, and task-focused communication. Two overarching dichotomies arose from our qualitative findings. First, Experiment Group participants were frustrated by the two-minute delay as it inhibited the natural pacing of their conversation. Meanwhile, the delay encouraged more intentional communication, where participants regularly and voluntarily exceeded the mandatory two-minute delay to carefully construct each message. Second, Experiment Group participants' ability to build social rapport with one another was disrupted by the intervention due to an increased time-pressure. Yet, these participants also appreciated an increased focus on the task, resulting in more efficient communication. Our quantitative analyses did not reveal statistically significant differences between groups with regards to the level of affect in participant messages and post-task scores in the Interpersonal Mindfulness Scale. Thus, while

we cannot conclude that the intervention had a universally positive effect, this work identified a potential for temporal design frictions to promote healthy online conflict conversations, and identified promising avenues for future research.

REFERENCES

[1] Mei Alonzo and Milam Aiken. 2004. Flaming in electronic communication. *Decision Support Systems* 36, 3 (Jan. 2004), 205–213. https://doi.org/10.1016/S0167-9236(02)00190-2

[2] Abdullah Alsaeedi and Mohammad Zubair. 2019. A Study on Sentiment Analysis Techniques of Twitter Data. *International Journal of Advanced Computer Science and Applications* 10, 2 (2019), 361–374. https://doi.org/10.14569/IJACSA.2019.0100248

[3] Monica Anderson and Brooke Auxier. 2020. 55% of U.S. social media users say they are 'worn out' by political posts and discussions. https://www.pewresearch.org/fact-tank/2020/08/19/55-of-u-s-social-media-users-say-they-are-worn-out-by-political-posts-and-discussions/

[4] Monica Anderson, Skye Toor, Lee Rainie, and Aaron Smith. 2018. Activism in the Social Media Age. https://www.pewresearch.org/internet/2018/07/11/activism-in-the-social-media-age/

[5] American Psychological Association. 2020. *Publication Manual of the American Psychological Association: 7th Edition.* APA, Washington, D.C.

[6] Robert M Bastress and Joseph D Harbaugh. 2004. Taking the Lawyer's Craft into Virtual Space: Computer-mediated interviewing, counseling, and negotiating. *Clinical Law Review* 10 (2004), 115–156.

[7] Amanda Baughan, Justin Petelka, Catherine Jaekyung Yoo, Jack Lo, Shiyue Wang, Amulya Paramasivam, Ashley Zhou, and Alexis Hiniker. 2021. Someone Is Wrong on the Internet: Having Hard Conversations in Online Spaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–22. https://doi.org/10.1145/3449230

[8] Joanne L. N. Bihari and Eugene G. Mullan. 2014. Relating Mindfully: A Qualitative Exploration of Changes in Relationships Through Mindfulness-Based Cognitive Therapy. *Mindfulness* 5, 1 (Feb. 2014), 46–59. https://doi.org/10.1007/s12671-012-0146-x

[9] Christopher C. Branch, Catherine I. Beaton, Michael McQuaid, and Elissa Weeden. 2021. Perceptions of Ethics in Persuasive User Interfaces. In *Persuasive Technology*, Raian Ali, Birgit Lugrin, and Fred Charles (Eds.). Vol. 12684. Springer International Publishing, Cham, 275–288. https://doi.org/10.1007/978-3-030-79460-6_22 Series Title: Lecture Notes in Computer Science.

[10] Kirk Warren Brown and Tim Kasser. 2005. Are Psychological and Ecological Well-being Compatible? The Role of Values, Mindfulness, and Lifestyle. *Social Indicators Research* 74, 2 (Nov. 2005), 349–368. https://doi.org/10.1007/s11205-004-8207-8

[11] Judee K. Burgoon, Charles R. Berger, and Vincent R. Waldron. 2000. Mindfulness and Interpersonal Communication. *Journal of Social Issues* 56, 1 (Jan. 2000), 105–127. https://doi.org/10.1111/0022-4537.00154

[12] S. Carpendale, S. Bardzell, M. Burnett, N. Kumar, and Madeline Balaam. 2018. Panel: Extending conversations about gender and HCI. In *Conference on Human Factors in Computing Systems - Proceedings :.* KTH, Media Technology and Interaction Design, MID, Montréal, QC, Canada, 1–6. https://doi.org/10.1145/3170427.3186325

[13] Jeff L Cochran. 2002. Empathetic communication for conflict resolution among children. *The Person-Centered Journal* 9, 2 (2002), 101–112.

[14] Kevin Coe, Kate Kenski, and Stephen A. Rains. 2014. Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication* 64, 4 (Aug. 2014), 658–679. https://doi.org/10.1111/jcom.12104

[15] Anna L. Cox, Sandy J.J. Gould, Marta E. Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, San Jose California USA, 1389–1397. https://doi.org/10.1145/2851581.2892410

[16] Debby Damen, Per Wijst, Marije Amelsvoort, and Emiel Krahmer. 2020. The Effect of Perspective-Taking on Trust and Understanding in Online and Face-to-Face Mediations. *Group Decision and Negotiation* 29 (Dec. 2020), 1121—-1156. https://doi.org/10.1007/s10726-020-09698-8

[17] Mathias Dekeyser, Filip Raes, Mia Leijssen, Sara Leysen, and David Dewulf. 2008. Mindfulness skills and interpersonal behaviour. *Personality and Individual Differences* 44, 5 (April 2008), 1235–1245. https://doi.org/10.1016/j.paid.2007.11.018

[18] Mehmet Engin Deniz, Seher Merve Erus, and Dilay Batum. 2020. Examining Marital Satisfaction in terms of Inter-personal Mindfulness and Perceived Problem Solving Skills in Marriage. *International Online Journal of Educational Sciences* 12, 2 (2020), 69–83. https://doi.org/10.15345/iojes.2020.02.005

[19] Paola Di Blasio and Luca Milani. 2008. Computer-mediated communication and persuasion: Peripheral vs. central route to opinion shift. *Computers in Human Behavior* 24, 3 (May 2008), 798–815. https://doi.org/10.1016/j.chb.2007.02.011

[20] Discord. 2021. Discord Support. https://support.discord.com/hc/en-us/articles/360016150952-Slowmode-Slllooowwwiiinng-down-your-channel#:~:text=To%20get%20this%20set%20up,can%20adjust%20the%20time%20intervals.

[21] Discord. 2022. About Discord | Our Mission and Values. https://discord.com/company

[22] Discord. 2022. DISBOARD: Public Discord Server List. https://disboard.org/

[23] William P Eveland and Myiah Hutchens Hively. 2009. Political Discussion Frequency, Network Size, and "Heterogeneity" of Discussion as Predictors of Political Knowledge and Participation. *Journal of Communication* 59, 2 (June 2009), 205–224. https://doi.org/10.1111/j.1460-2466.2009.01412.x

[24] Jenny Fan and Amy X. Zhang. 2020. Digital Juries: A Civics-Oriented Approach to Platform Governance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376293

[25] Frida Kerner Furman. 2010. Compassionate listening as a path to conflict resolution. *STUDY OF PEACE AND CONFLICT* 2009 (2010), 24–38.

[26] Lawrence H. Ganong, Marilyn Coleman, Richard Feistman, Tyler Jamison, and Melinda Stafford Markham. 2012. Communication Technology and Postdivorce Coparenting. *Family Relations* 61, 3 (July 2012), 397–409. https://doi.org/10.1111/j.1741-3729.2012.00706.x

[27] James J Gibson. 1977. The Theory of Affordances. In *Perceiving, Acting, and Knowing: Toward and Ecological Psychology*. John Wiley & Sons, Inc., New York. https://monoskop.org/images/2/2c/Gibson_James_J_1977_The_Theory_of_Affordances.pdf

[28] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, New Haven.

[29] Sandy J. J. Gould, Lewis L. Chuang, Ioanna Iacovides, Diego Garaialde, Marta E. Cecchinato, Benjamin R. Cowan, and Anna L. Cox. 2021. A Special Interest Group on Designed and Engineered Friction in Interaction. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–4. https://doi.org/10.1145/3411763.3450404

[30] Joseph Graf, Joseph Erba, and Ren-Whei Harn. 2017. The Role of Civility and Anonymity on Perceptions of Online Comments. *Mass Communication and Society* 20, 4 (July 2017), 526–549. https://doi.org/10.1080/15205436.2016.1274763

[31] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 1–35. https://doi.org/10.1145/3479610 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[32] Gunnar Harboe and Elaine M. Huang. 2015. Real-World Affinity Diagramming Practices: Bridging the Paper-Digital Gap. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 95–104. https://doi.org/10.1145/2702123.2702561

[33] Claire Hardaker. 2010. Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research. Language, Behaviour, Culture* 6, 2 (Jan. 2010), 215–242. https://doi.org/10.1515/jplr.2010.011

[34] Danula Hettiachchi and Jorge Goncalves. 2019. Towards Effective Crowd-Powered Online Content Moderation. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction (OZCHI'19)*. Association for Computing Machinery, New York, NY, USA, 342–346. https://doi.org/10.1145/3369457.3369491

[35] Myiah J Hutchens, Vincent J Cicchirillo, and Jay D Hmielowski. 2015. How could you think that?!?!: Understanding intentions to engage in political flaming. *New Media & Society* 17, 8 (Sept. 2015), 1201–1219. https://doi.org/10.1177/1461444814522947

[36] Bruce Hyde and Jeffery L. Bineham. 2000. From debate to dialogue: Toward a pedagogy of nonpolarized public discourse. *Southern Communication Journal* 65, 2-3 (March 2000), 208–223. https://doi.org/10.1080/10417940009373168

[37] Kokil Jaidka, Alvin Zhou, and Yphtach Lelkes. 2019. Brevity Is the Soul of Twitter: The Constraint Affordance and Political Discussion. *Journal of Communication* 69, 4 (2019), 29.

[38] Sirkka L. Jarvenpaa and Dorothy E. Leidner. 1998. Communication and Trust in Global Virtual Teams. *Journal of Computer-Mediated Communication* 3, 6 (June 1998), 791–815. https://doi.org/10.1111/j.1083-6101.1998.tb00080.x

[39] Jason Jay and Gabriel Grant. 2017. *Breaking through gridlock: The power of conversation in a polarized world*. Berrett-Koehler Publishers, Oakland, CA.

[40] Jon Kabat-Zinn. 1993. Mindfulness meditation: Health benefits of an ancient Buddhist practice. In *Mind/Body Medicine*. Consumer Reports, Yonkers, NY, 259–275.

[41] Jon Kabat-Zinn. 2013. *Full Catastrophe Living* (revised and updated edition ed.). Bantam Dell Publishing Group, New York, NY.

[42] Yoram M. Kalman, Lauren E. Scissors, Alastair J. Gill, and Darren Gergle. 2013. Online chronemics convey social information. *Computers in Human Behavior* 29, 3 (May 2013), 1260–1269. https://doi.org/10.1016/j.chb.2012.12.036

[43] Lena Kühn, Charlotte Hermann, and Laurens Boer. 2019. Design Research Themes for Mindful Interaction. In *Proceedings of the 17th Scandinavian Conference on Health Informatics*. Linköping University Electronic Press, Oslo, Norway, 24–29.

[44] Noam Lapidot-Lefler and Azy Barak. 2012. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior* 28, 2 (2012), 434–443. https://doi.org/10.1016/j.chb.2011.10.014 Place: Netherlands Publisher: Elsevier Science.

[45] Heidemarie K. Laurent, Robin Hertz, Benjamin Nelson, and Sean M. Laurent. 2016. Mindfulness during romantic conflict moderates the impact of negative partner behaviors on cortisol responses. *Hormones and Behavior* 79 (March 2016), 45–51. https://doi.org/10.1016/j.yhbeh.2016.01.005

[46] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–15. https://doi.org/10.1145/3313831.3376672

[47] Melinda Stafford Markham, Jaimee L. Hartenstein, Yolanda T. Mitchell, and Ghadir Aljayyousi-Khalil. 2017. Communication Among Parents Who Share Physical Custody After Divorce or Separation. *Journal of Family Issues* 38, 10 (July 2017), 1414–1442. https://doi.org/10.1177/0192513X15616848

[48] Adrienne Massanari. 2017. #Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. *New Media & Society* 19, 3 (March 2017), 329–346. https://doi.org/10.1177/1461444815608807

[49] Joseph Edward McGrath. 1984. *Groups: Interaction and performance.* Prentice-Hall, Englewood Cliffs, N.J.

[50] Brian E Mennecke, Joseph S Valacich, and Bradley C Wheeler. 2000. The Effects of Media and Task on User Performance: A Test of the Task-Media Fit Hypothesis. *Group Decision and Negotiation* 9 (2000), 507–529.

[51] Rocío Galarza Molina and Freddie J. Jennings. 2018. The Role of Civility and Metacommunication in Facebook Discussions. *Communication Studies* 69, 1 (Jan. 2018), 42–66. https://doi.org/10.1080/10510974.2017.1397038

[52] Luke Munn. 2020. Angry by Design: Toxic Communication and Technical Architectures. *Humanities and Social Sciences Communications* 7, 1 (July 2020), 1–11. https://doi.org/10.1057/s41599-020-00550-7

[53] Diana C. Mutz. 2002. Cross-cutting Social Networks: Testing Democratic Theory in Practice. *American Political Science Review* 96, 1 (March 2002), 111–126. https://doi.org/10.1017/S0003055402004264

[54] Chikashi Nobata, Joel Tetreault, Achint Thomas, Yashar Mehdad, and Yi Chang. 2016. Abusive Language Detection in Online User Content. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Montréal Québec Canada, 145–153. https://doi.org/10.1145/2872427.2883062

[55] Donald A. Norman. 1999. Affordance, Conventions, and Design. *interactions* 6, 3 (1999), 38–43. http://dl.acm.org/citation.cfm?id=301168

[56] James W. Pennebaker, RJ Booth, Ryan L. Boyd, and ME Francis. 2021. LIWC | Linguistic Inquiry and Word Count. http://liwc.wpengine.com/

[57] Simon T. Perrault and Weiyu Zhang. 2019. Effects of Moderation and Opinion Heterogeneity on Attitude towards the Online Deliberation Experience. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300247

[58] Steven D. Pratscher, Phillip K. Wood, Laura A. King, and B. Ann Bettencourt. 2019. Interpersonal Mindfulness: Scale Development and Initial Construct Validation. *Mindfulness* 10, 6 (June 2019), 1044–1061. https://doi.org/10.1007/s12671-018-1057-2

[59] Manton Reece. 2018. Manton Reece - Cutting Edge to Old Fashioned. Retrieved {2021-11-19} from https://www.manton.org/2018/10/29/cutting-edge-to.html

[60] Carla Anne Roos, Namkje Koudenburg, and Tom Postmes. 2021. Dealing with disagreement: The depolarizing effects of everyday diplomatic skills face-to-face and online. *New Media & Society* 24, 9 (March 2021), 146144482199304. https://doi.org/10.1177/1461444821993042

[61] Johnny Saldaña. 2013. *The coding manual for qualitative researchers* (2nd ed ed.). SAGE, Los Angeles.

[62] Amy J Schmitz. 2019. Expanding Access to Remedies through E-Court Initiatives. *Buffalo Law Review 89* 67, 1 (2019), 89–164.

[63] Raymond Scupin. 1997. The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology. *Human Organization* 56, 2 (1997), 233–237. https://doi.org/10.17730/humo.56.2.x335923511444655

[64] Joseph Seering. 2020. Reconsidering Self-Moderation: the Role of Research in Supporting Community-Based Models for Online Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–28. https://doi.org/10.1145/3415178

[65] Bruce M. Smyth, Giverney Ainscough, and Jason L. Payne. 2020. Modes of Communication between High-Conflict Separated Parents: Exploring the Role of Media Multiplexity and Modality Switching. *Journal of Family Communication* 20, 3 (July 2020), 189–205. https://doi.org/10.1080/15267431.2020.1754826

[66] Katherine W Song, Janaki Vivrekar, Lynn Yeom, Eric Paulos, and Niloufar Salehi. 2021. Crank That Feed: A Physical Intervention for Active Twitter Users. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–6. https://doi.org/10.1145/3411763.3451817

[67] Kumar Bhargav Srinivasan, Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Chenhao Tan. 2019. Content Removal as a Moderation Strategy: Compliance and Other Outcomes in the ChangeMyView Community. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 163:1–163:21. https://doi.org/10.1145/3359265

[68] Miriah Steiger, Timir J Bharucha, Sukrit Venkatagiri, Martin J. Riedl, and Matthew Lease. 2021. The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445092

[69] Andie Storozuk, Marilyn Ashley, Véronic Delage, and Erin A. Maloney. 2020. Got Bots? Practical Recommendations to Protect Online Survey Data from Bot Attacks. *The Quantitative Methods for Psychology* 16, 5 (May 2020), 472–481. https://doi.org/10.20982/tqmp.16.5.p472

[70] Susan G. Straus. 1999. Testing a Typology of Tasks: An Empirical Validation of McGrath's (1984) Group Task Circumplex. *Small Group Research* 30, 2 (April 1999), 166–187. https://doi.org/10.1177/104649649903000202

[71] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Montréal Québec Canada, 613–624. https://doi.org/10.1145/2872427.2883081

[72] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (March 2010), 24–54. https://doi.org/10.1177/0261927X09351676

[73] Nađa Terzimehić, Renate Häuslschmid, Heinrich Hussmann, and m.c. schraefel. 2019. A Review & Analysis of Mindfulness Research in HCI: Framing Current Lines of Research and Future Opportunities. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300687

[74] Milo Trujillo, Maurıcio Gruppi, Cody Buntain, and Benjamin D Horne. 2020. Hateful YouTube: Characterizing the Social Video-hosting Platform BitChute. In *6th International Conference on Computational Social Science IC2S2*. MIT Media Laboratory, Boston, MA USA, 3.

[75] Devangi Vivrekar. 2018. *Persuasive Design Techniques in the Attention Economy: User Awareness, Theory, and Ethics*. Ph.D. Dissertation. Stanford University.

[76] Joseph B. Walther. 1992. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research* 19, 1 (1992), 52–90. https://doi.org/10.1177/009365092019001003 _eprint: https://doi.org/10.1177/009365092019001003.

[77] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A field trial of privacy nudges for facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 2367–2376. https://doi.org/10.1145/2556288.2557413

[78] Chris Wells, Katherine J. Cramer, Michael W. Wagner, German Alvarez, Lewis A. Friedland, Dhavan V. Shah, Leticia Bode, Stephanie Edgerly, Itay Gabay, and Charles Franklin. 2017. When We Stop Talking Politics: The Maintenance and Closing of Conversation in Contentious Times: Political Talk in Contentious Times. *Journal of Communication* 67, 1 (Feb. 2017), 131–157. https://doi.org/10.1111/jcom.12280

[79] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (Nov. 2018), 4366–4383. https://doi.org/10.1177/1461444818773059

[80] Mary-Frances Winters. 2017. *We can't talk about that at work!: How to talk about race, religion, politics, and other polarizing topics*. Berrett-Koehler Publishers, Oakland, CA.

[81] Yan Xia, Haiyi Zhu, Tun Lu, Peng Zhang, and Ning Gu. 2020. Exploring Antecedents and Consequences of Toxicity in Online Discussions: A Case Study on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 108:1–108:23. https://doi.org/10.1145/3415179

[82] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. 2018. What is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber. In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18*. ACM Press, Lyon, France, 1007–1014. https://doi.org/10.1145/3184558.3191531