

Someone Is Wrong on the Internet*: Having Hard Conversations in Online Spaces

AMANDA BAUGHAN, Computer Science & Engineering, University of Washington, USA

JUSTIN PETELKA, Information School, University of Washington, USA

CATHERINE J. YOO, University of Washington, USA

JACK LO, University of Washington, USA

SHIYUE WANG, University of Washington, USA

AMULYA PARAMASIVAM, University of Washington, USA

ASHLEY ZHOU, University of Washington, USA

ALEXIS HINIKER, Information School, University of Washington, USA

Good faith disagreements and healthy conflict management are essential to deliberative democracy and building strong relationships. People increasingly use computer-mediated communication during disagreements, which raises the question of how technology and design impact users' disagreements and relationships. We conducted a mixed-methods study with 257 total participants to understand how design impacts disagreements across both existing social media platforms and novel, user-generated designs. Through interviews, a survey, and storyboard evaluations, we found that users often want to discuss challenging topics online but avoid them due to fear of hurting their relationships. Further, we found that users are most excited about design interventions that empower collective group action, humanize others online, or support channel switching to more private or socially rich contexts. Our results suggest that although technology has the potential to support users during conflict, it is also rife with possibilities to do more harm than good by diluting users' intentions, intruding, or backfiring. We introduce "interpersonal design," which centers relationships in the design process, an essential step in supporting users in the challenging task of arguing well.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**.

Additional Key Words and Phrases: computer-mediated communication; online arguments; social media design; conflict; interpersonal design

ACM Reference Format:

Amanda Baughan, Justin Petelka, Catherine J. 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. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 156 (April 2021), 22 pages. <https://doi.org/10.1145/3449230>

*<https://xkcd.com/386/>

Authors' addresses: Amanda Baughan, baughan@cs.washington.edu, Computer Science & Engineering, University of Washington, USA; Justin Petelka, Information School, University of Washington, USA; Catherine J. Yoo, University of Washington, USA; Jack Lo, University of Washington, USA; Shiyue Wang, University of Washington, USA; Amulya Paramasivam, University of Washington, USA; Ashley Zhou, University of Washington, USA; Alexis Hiniker, alexisr@uw.edu, Information School, University of Washington, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2021/4-ART156 \$15.00

<https://doi.org/10.1145/3449230>

1 INTRODUCTION

An enormous amount of communication now occurs online, as the number of people on social media platforms is higher than ever [1]. Unsurprisingly, online disagreements are common; more than 20% of people say that, at times, they find themselves embroiled in conflict online [20]. Many people try to avoid online arguments because of their tendency to turn toxic [18, 44] and damage relationships [47, 55]. However, conflicts serve an important role in close interpersonal relationships, as they allow people to consider their needs and create meaningful changes [53]. Arguments are also a useful mechanism for exchanging and testing ideas, and ideological arguments are essential to deliberative democracy [21, 45]. Prior work has shown that when people argue, they often want to leverage digital platforms [38], and it is possible, but rare, to argue productively online [51]. This suggests that it is a worthwhile endeavor to design platforms that support engaging in constructive disagreements, rather than avoiding disagreement altogether.

Therefore, we investigated users' experiences arguing with strangers, friends, and loved ones across common social and communication platforms, investigating how cross-cutting designs and platform-specific features relate to the tenor of conversations online. Specifically, we ask:

- **RQ1:** Given the many forms of technology-mediated communication people engage in, which features and affordances do users perceive as most relevant to their arguments online?
- **RQ2:** Which cross-cutting design approaches would support constructive online arguments, and what are the perceived risks and benefits of those approaches?

To examine these questions, we conducted a three-part, mixed-methods study. We first conducted semi-structured interviews with 22 people, asking about their experiences arguing on social media platforms and prompting them to generate novel design ideas to support arguing online. We then administered a survey to 137 participants to examine and quantify interview themes across a broader sample. Finally, we conducted an evaluation of 12 design approaches distilled from interviewees' design ideas, with 98 additional participants.

We contribute an overview of the features and affordances of ten popular social platforms that are relevant to users in the context of online arguments, along with their goals for disagreements online. We find that people have a desire to discuss challenging topics on social media they currently use, but they often hold back due to fear of an argument. When in conflict, users highly value privacy and the ability to switch from public to private channels. Users also value interventions that empower them and their communities to take action to shape their online spaces, as well as interventions that remind them of the humanity of others online. Across novel design evaluations, we find that users are excited about the potential for technology to support their relationships, but there are also many ways in which design can make a difficult situation worse. Users are categorically opposed to design interventions which they perceive to be too intrusive and heavy-handed, or those that would dilute the intent of their communication.

Therefore, we introduce the concept of *interpersonal design* that foregrounds users' relationships in the design process. We include several suggestions for practitioners to incorporate an interpersonal design approach, such as incorporating research with pairs of participants and breaking down research into phases of conflict, such as initiating and resolving arguments. We hope that this work will encourage designers and researchers to consider how to support authentic, good-faith online arguments that empower users. Awareness of the nuances of this challenging design space can help designers and researchers support users' goal of engaging in hard conversations online constructively to strengthen their relationships and communities.

2 RELATED WORK

A large body of prior work has investigated online arguments, including how social behavior changes online [24, 33, 36, 49, 57], how social media design affects online arguments [18, 27, 37, 45, 58], and how various interventions can facilitate productive online discussions [4, 6, 7, 10, 29, 30, 39, 48, 52]. Here, we describe the substantial foundation to which we contribute.

2.1 Characteristics and Consequences of Online Arguments

A number of studies demonstrate systematic differences between online and offline conversations [24, 33, 36, 49, 57]. Nuances of body language and tone—only present during real-time, face-to-face (or voice-to-voice) communication—convey vast amounts of information to conversational partners [40]. Without this information, users find less common ground with others [36] and become more uninhibited [49, 57], making them more likely to instigate heated debates but no more likely to resolve them. At its worst, online disinhibition allows toxic behaviors to flourish, which can include incivility [8], harassment [6, 11, 26], trolling [10], and cyberbullying [24, 32]. Toxic online disinhibition may manifest from a desire to place blame elsewhere for poor performance in a game [32] or due to heightened social anxiety and lack of self-regulation [42]. Specific design affordances such as invisibility, perceived anonymity [49, 57], and lack of eye-contact [33] contribute to toxic disinhibition and decreased empathy online. Although some work has found that certain users are more prone to incivility online than others [39], other work has found that “anyone can become a troll” [10].

The impacts of toxic online disinhibition can be long-lasting and affect both online and offline relationships. Disagreeing with someone’s posted content is a common motivation for unfriending on Facebook [28, 45, 47] and users often regret posting angry content because of the damage done to self-presentation and relationships [55]. Online posts can also lead to more severe consequences such as job loss or the dissolution of romantic relationships [55]. However, not all online arguments end in lasting damage. Sometimes, users report maintaining connections that they have considered severing because they value the relationship [31]. They also sometimes consider an inflammatory online post to be a minor issue in a strong friendship, with weak ties being the most vulnerable to long-lasting repercussions [18]. Online disinhibition can also manifest as *benign disinhibition*, in which people offer more kindness and generosity to others online than they do in person [49]. This suggests that there is room for designers to capitalize on benign disinhibition and nudge users toward patterns of dialogue that enable them to remain open-minded in the face of online conflict.

2.2 Differences in Online Arguments across Different Popular Platforms

Research has found that arguments on Facebook [18], Twitter [58], and Reddit [27], often create strain in relationships and do not result in changed opinions. For instance, Liu and Weber [37] found that Twitter is limited in its ability to foster effective, democratic discourse, and the effect of Twitter arguments is typically to reinforce pre-existing views [58]. Users also consider tweets to lack the contextual depth necessary for engaging in constructive dialogue [45]. Facebook is considered a difficult place to maintain friendships with those who have different opinions, especially during heated political debates [18]. However, when users have an explicit goal of sharing and considering opposing viewpoints, as found in the subreddit ChangeMyView, they often find engaging in the arguments to be a positive and constructive experience. Still, in the majority of instances, users do not change their view [27, 51]. This highlights the important role of people’s attitudes when arguing online. Those who are open to or actively seek out diverse opinions are more likely to engage respectfully and less likely to experience polarization [46]. Liao and Fu [34] found that the tone of an argument affects participants’ likelihood of changing their views, and people are less likely to

engage with content that contradicts their views if they perceive the interaction to be threatening. Additionally, polite disagreements on ChangeMyView seemed to be most successful [27]. Thus polite and respectful conversational tone may promote engagement on controversial topics.

These studies provide evidence for how online arguments can still be a positive experience, even when no one changes their opinion. It also shows how users have different experiences in different online spaces. People sometimes channel switch across social media platforms to tailor their platform to their intended audience and content [59] and to have greater freedom of self-expression [19]. Hence, it is important to holistically evaluate all social media people use in the context of online arguments when evaluating the impact of design on online conflict.

2.3 Novel Designs for Constructive Discourse

A number of studies illustrate how interface design plays a differential role in online discourse. People examine the environment around them for signals to help determine appropriate behavior, such as speaking more quietly as a consequence of being in a library [2]. This kind of normative influence on social behaviors also occurs online [48] and can affect behaviors for better [48] or worse [10]. Sukumaran et al. [48] found that certain online environmental cues, such as language and design, can affect the amount of thoughtfulness people perceive in others' statements and the amount of thoughtfulness they exhibit in their own communication. Website design creates an almost instantaneous affective response [35], which highlights the value of designing for constructive conversation. Thus, although online environments can strip away important non-verbal cues, design choices can compensate for some of that deficit and systematically influence the extent to which users engage in prosocial communication patterns online.

Given the evidence that design can prompt behavioral changes [48], a number of prior works have invented and evaluated novel designs to improve discourse. A series of work by Kriplean et. al. has illustrated the benefits of visualizing others' opinions in a nuanced way [29] and of engaging in active "listening" by repeating, clarifying, and voting on interpretations of people's posts [30]. Participants reported that discourse on these platforms was more productive than in other online spaces they occupy [29, 30]. These studies show that designers can systematically influence the way in which users engage in conversation and respond to conflict. Our work leverages these findings and extends them by examining existing design patterns that influence arguments and novel design approaches that could be applied across contexts to promote more constructive arguments online.

3 METHODS

Here we outline the interviews, survey, and storyboard evaluation that comprised our investigation of how design affects online arguments.

3.1 Interviews

3.1.1 Participants. We recruited 22 adults to participate in a two-part interview study. Participants were a convenience sample recruited through snowball sampling and posts on email lists and social media. The plurality of interviewees were college students ($n = 12$). 68% of participants ($n = 15$) provided their age, gender, and race as part of a free-response form, of which four were women, ten were men, and one was non-binary. One of the men was trans; no other participants reported being trans or cisgender. The average age was 22.4 ($sd = 4.8$). The plurality of participants self-identified as Asian or Asian American ($n = 12$). During recruiting, the research team explained that we were studying online arguments, and that participants were only eligible if they had argued with someone online at least once in the past. Interviews lasted approximately 30-45 minutes, and each participant received a US\$20 gift certificate to Amazon.

3.1.2 Materials and Procedures. The interview was conducted in two parts. First, we asked participants to name all of the social media and messaging apps and websites they use at least once a week. We repeated the same series of questions about their experiences with arguments for each of these platforms, asking things like, “*Do you feel like you can talk about challenging topics with other people when you’re using [platform]?*” We then asked if the participant could recall a specific argument in which they had engaged on the platform of interest, and if so, to describe the details of this event. Second, participants took part in a sketching exercise in which they were asked to invent three different ways that, “*apps or websites might help people have more constructive conversations online.*” The participant then walked the researcher through each of the design ideas, and the researcher asked follow-up questions to encourage the participant to elaborate as needed.

3.1.3 Analysis. Interviews were transcribed in their entirety, anonymized, and analyzed via an inductive process [12]. To analyze part one of the interview, six members of the research team read independent subsets of interview transcripts to identify emergent themes through open coding. The team met weekly for one month to iteratively discuss and refine these themes and compare examples from different participants. Each team member then selected a small number of themes to investigate more thoroughly, and over the course of one month, read all interview transcripts to identify instances of the selected themes. These quotes were used to collaboratively create an affinity diagram using Miro. The team then examined and refined each cluster as a group. One researcher then selected and organized key quotes based on the affinity diagram, and an additional two researchers who had not participated in the initial analysis reviewed these quotes to diversify perspectives and broaden and strengthen codes. The details of the design sketching analysis are discussed in more detail in Section 3.3.

3.2 Survey

To quantitatively investigate the themes that surfaced in interviews across a broader user base, we conducted a survey with an additional 137 participants who self-reported engaging at least once per week with at least one of the ten social or communication platforms that surfaced repeatedly in interviews. These platforms are listed in Table 1.

3.2.1 Participants. The 137 survey participants ranged from 18 to 64 years old ($M=33.9$, $SD=9.8$). All participants reported in a free-response gender prompt that they self-identify as men ($n=82$) or women ($n = 55$). Six unique racial groups were represented, the majority of which were White ($n=88$) and Asian or Pacific Islander ($n=25$). The rest ($n=24$) of the participants identified as Black, Hispanic/Latino, Native American/Alaska Native, or a combination of races and ethnicities. Participants’ education ranged from high school degree or equivalent ($n=20$) to higher education such as a master’s, doctorate, or professional degree ($n=16$), with the plurality of participants having achieved a bachelor’s degree ($n=64$). Participants’ political leanings ranged from extremely conservative ($n=4$) to extremely liberal ($n=17$), with the plurality self-identifying as moderate ($n=30$). Most participants’ annual income was in the range of US \$25,000-50,000 ($n=50$).

3.2.2 Materials and Procedures. We deployed the survey online on Amazon’s Mechanical Turk platform. The survey was composed of several subsections:

- **General Usage.** Participants first reported which social media and messaging platforms they use on a weekly basis. For each of these platforms, participants were then asked how many hours per week they spend using it and whether they could remember having had an argument on that platform with someone they knew personally.

| | Text-forward | Image-forward |
|---------|--|--------------------------------|
| Public | Facebook Twitter Reddit | YouTube Tumblr Instagram |
| Private | WhatsApp Messenger (Facebook) WeChat | Snapchat |

Table 1. Social media platforms categorized by participants' perceptions. Participants compared and contrasted perceived public/private and text/image-forward platforms in interviews without prompting.

- *Specific Argument Instance.* Participants were then asked to describe the most recent argument they could recall having online, including: 1) the platform where the argument occurred, 2) what the argument was about, and 3) who they argued with.
- *Per Platform Experiences.* For each platform the participant reported using at least once per week, we asked participants a series of Likert-style questions probing whether they argue on the platform often and whether they feel they can discuss challenging or controversial topics. We also asked open-ended questions about their experience using the platform.
- *Demographics.* The survey concluded with basic demographic questions.

Participants were compensated US\$3 for completing the survey, which took 16 minutes on average.

3.2.3 Analysis. We extracted two datasets from participants' survey responses. We used participants' descriptions of the most recent argument they had online to generate a set of 137 unique arguments, one per participant. We coded these arguments for the platform where they occurred, which was listed explicitly by participants in 133 cases. We also generated a long-form dataset with one entry for each platform that the participant said they used at least once per week, leading to a variable number of data points per participant, and a total of 579 total entries. Each entry included the platform, the participant, and the participant's responses to questions about that platform. As shown in Table 1, platforms were categorized according to themes that emerged during interviews, specifically, the differences participants described in their experiences on social media as a function of 1) audience, and 2) the perceived salience of text and images. Two researchers conducted inductive analysis and affinity diagramming on open responses to the prompt "*features that make arguing on [platform] easier/harder.*" We ran a series of statistical analyses which were guided by the themes generated in interviews, as the purpose of the survey was to triangulate interview findings with a more diverse population.

3.3 Storyboard Evaluation

Finally, we analyzed a set of novel designs, grounded in the design ideas participants described during interviews, to investigate how platforms can better support constructive arguments.

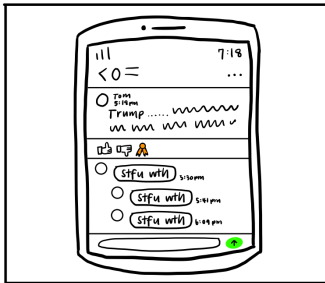
3.3.1 Materials and Procedures. From the interviews' design sketching portion, the research team constructed a dataset composed of all participants' design ideas. These were then printed on physical cards, with one design concept per card, and the research team clustered the design ideas into a physical affinity diagram. Six members of the research team thematically analyzed these design ideas and met several times to iteratively discuss and refine these emergent themes. A final set of 12 themes were chosen, and each of the six members was assigned two themes and independently developed storyboards for those themes. Many of these 12 storyboards encompassed multiple unique design interventions within the theme. Final storyboards were decided by consensus through discussing and combining ideas within each design approach. An example storyboard is

shown in Figure 1, and all storyboards are made available with the supplemental materials. We embedded all 12 storyboards in a new survey, which we deployed on Amazon’s Mechanical Turk. We asked participants how often they used social media and messaging apps, and which specific platforms they used. We then presented each of the 12 storyboards sequentially and in random order. For each storyboard, participants responded to two scaled-response questions asking them to assess their reaction to the concept and how willing they were to try it, from “Very Negative/Unwilling to Try” to “Very Positive/Willing to Try.” There was also a required open-ended question asking what they liked and disliked about each storyboard. The survey concluded with demographic questions.

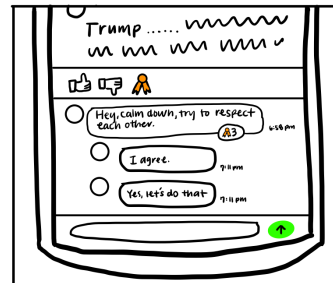
3.3.2 Participants. A total of 98 new participants each evaluated all 12 storyboards. We collected data from 58 men, 39 women, and 1 agender person. Participants ranged from 22 to 65 years old ($M=34.7$, $SD=9.0$). The majority of participants were White ($n=78$), and the rest of the participants identified as Black, Asian or Pacific Islander, Native American, or a combination of different racial identities. Participants’ education ranged from high school diploma or less ($n=17$) to postgraduate degree ($n=8$). A plurality of participants had a bachelor’s degree ($n=37$).

3.3.3 Analysis. We performed quantitative within-subject analyses comparing participants’ reactions across storyboards. To identify themes in aspects of the designs that participants valued and disliked, we used an inductive process to code the 1,176 open-ended responses. To do so,

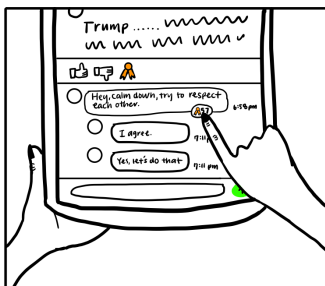
Democracy



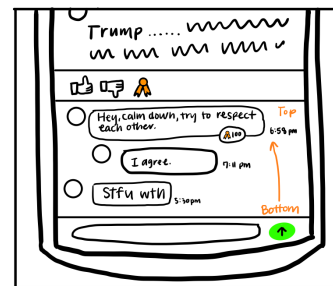
Jack reads through comments on a heated political post and sees nothing but chaos.



Jack finds a constructive comment buried at the bottom and decides to use the "constructive" button on it.



Other people proceed to do the same and gradually the comment gains visibility.



So the comment moves from the bottom to the top and reduces the tension of the post.

Fig. 1. An example storyboard. In this storyboard, the platform enables the user to tag a comment as “constructive” in a conversational thread online, which results in elevating the comment

six members of the research team first independently coded the same small subset of responses, discussing potential themes as a group. Over seven weeks, the team met weekly, iteratively coding subsets of the data and refining codes, and, after finalizing all codes, coding the entire dataset. At least two coders coded each response in this final pass.

3.4 Limitations of Methodology

It is essential to note that this was not a hypothesis-driven investigation, and although we triangulate our qualitative findings with quantitative analyses, including inferential comparisons, we did not begin this work with strong *a priori* hypotheses. We mitigate this limitation somewhat by reporting on effects that were robust within our dataset and by choosing analyses guided by our interview findings. This work was also conducted in a Western context, and the values and reactions that our participants expressed may not translate well to other cultures and locations. This is reflected in the social media platforms that participants discussed, as primarily Western social media platforms informed the interview and survey data. In our design evaluation, we only collected speculative data about storyboards; we did not give participants the opportunity to experience any of these design concepts *in situ*. Implementing and evaluating these designs is an important next step for future works.

4 RESULTS: IMPACT OF DESIGN ON ONLINE ARGUMENTS

In interviews and survey responses, participants explained that the design of an online space can support having challenging conversations constructively and in ways that could not occur face-to-face, especially when initiating conversations. However, they simultaneously described other common properties of online spaces that interfere with the delicate process of working through a challenging topic and strip away important cues that occur in face-to-face contexts. Here, we describe these design tensions. Quotes have been lightly edited for readability and are denoted with (P) for interviews and (S) for surveys.

4.1 Hunger for Hard Conversations

Many participants told us that they “*try to avoid*” (P4) online arguments and “*try to stay somewhat hidden*” (P1) in social spaces online. Many perceived online arguments to be futile, and they favored talking in person or avoiding difficult topics altogether. P2 lamented that, “*it’s hard to open a conversation where you’re almost not allowed to have a conversation... it’s just too much energy.*” However, despite the fact that participants described working to avoid controversial subjects, they also explained that they *wished* they could discuss these topics in online spaces. As P12 explained, “*I wish I could... talk to more people about my political stance, especially family. But I feel like if I were to do that, it would lead to arguments.*” Similarly, P22 described a desire for more, “*nuanced conversations on identity,*” saying that, “*there are sometimes where I do want to engage in that kind of conversation; sort of like academic almost conversation.*” However the participant further explained that they do not discuss identity online because, “*it just turns into, ‘You don’t agree with me. You’re terrible. Blocked.’ So, yeah, it’s pretty unproductive.*”

Participants also stated that they want to work through difficult conversations more often, because they provide important growth opportunities. P9 explained, “*I pushed myself to talk about it, and it makes me learn more about relationships. It’s really important to me and to my relationship with my friends.*” Across all interviewees, 68% ($n=15$) said they regularly use at least one app or platform where they want to discuss topics that they currently avoid, suggesting a hunger for conversations. Politics was the most frequently cited topic that participants said they want to discuss but avoid, along with ethics, religion, race, identity, and personal details about their life. This suggests that designs to support these conversations would be valued.

4.2 Anonymity and Ephemerality Alleviate Fear and Avoidance

In instances where participants did initiate difficult conversations online, they described how anonymity and ephemerality alleviated some of their fears. Participants said talking in public spaces online is made easier because of *“the fact that I don’t use my full name,”* (S127) explaining that *“if you use a name that isn’t linked to your real identity, you can usually discuss topics without having things connected to your real self”* (S94). However, anonymity did not fully relieve participants’ fears when discussing something controversial. This was particularly salient for public social media platforms, where participants were afraid of harassment from both within and outside of the platform. *“It’s not that difficult to connect someone’s Twitter name to their real identity. If you have a controversial view and someone finds out who you really are, you might suffer harassment outside of Twitter”* (S94). Participants echoed this sentiment for YouTube, saying that *“when you join a social network, whatever it is, you’re exposed and you have to watch the things you write, because they can be used against you”* (S111). Anonymity also resulted in defeatist attitudes about online arguments, because *“it gives me the feeling that I am arguing over nothing with nobodies [on Reddit]”* (S7).

A similar theme developed in users’ discussions of ephemerality on Snapchat. Participants said that the disappearing images and chat discussions made it *“the safest place”* (S25) to have difficult discussions. However, because *“conversations disappear after a certain amount of time, so there is no way to go back and look at what was said”* (S48), it also made arguing more difficult. *“It’s nice to be able to go back and point out specific sentences sometimes, and you can’t do that with Snapchat”* (S102). Users also stated how even on Snapchat *“someone could always share [a] screenshot. Anything posted or written online is never 100% secure”* (S136). Thus, participants indicated that while anonymity and ephemerality alleviate fear and risk when discussing controversial topics, they are left unsatisfied with their exposure to harassment and conversations being taken out of context. This demonstrates how these features are not sufficient for facilitating productive arguments online.

4.3 Asynchronicity Facilitates Thoughtfulness

Many participants described the time dilation (asynchronicity) of online conversations as a mechanism for encouraging constructive dialogue. For example, P6 explained:

“Online I have more time, so I can think out my responses. So I can actually have a full-blown argument that lasts longer than a couple minutes... And the person I’m arguing with will have time to come up with what to say.”

Others concurred, saying that online, *“it’s easier to step back and take a minute to think. In real life you can’t do that. There’s a lot more emotional charge”* (P18). However, participants also said that asynchronicity can make communication harder, because, *“sometimes it may be a while before the other person sees the message,”* (S63) and, *“you may write something controversial and nobody will reply to it”* (S135). The latter was particularly salient for YouTube and Instagram comments.

Conversely, some participants thought messaging apps such as Messenger and WhatsApp could facilitate real-time conversations. *“The conversations happen in real time. It is like a real conversation or exchanging text messages. Given that, you can have a healthy back and forth with another person”* (S80). However, they also said that, *“instant messaging makes for less in depth arguments”* (S30). This feedback on time dilation points to seemingly conflicting desires for conversations to feel natural and as though they occur in real-time, while also affording more careful consideration of wording than in-person conversations.

4.4 Privacy and Channel Switching Facilitate Authenticity

Participants described self-censoring to project a certain image of themselves on social media, consistent with prior work examining how people cultivate their online persona [5]. Participants

explained that they are more authentic when using private or audience-restricted platforms, making it easier to bring up or respond to controversial topics.

“[On a] Twitter or Facebook newsfeed, you don’t tend to be honest because other people are looking at you, and sometimes you don’t admit your failures because other people are looking at your stuff. So to be honest and then have a honest conversation, I think it’s better to have something that’s private.” (P1)

Thus, participants explained that as the audience for their communication shrinks, the likelihood of achieving shared understanding increases. As a result, users seek out platforms that provide more privacy when they want to discuss something challenging, and they refrain from responding to controversial content when they are aware of having an audience. Other participants contrasted experiences with larger and smaller audiences, saying things like, *“I can talk about it [sensitive topics] on Messenger, but I don’t want to comment on it and all the people see [on Facebook].” (P9).*

Additionally, participants said that features of messaging apps such as *“seeing the online status of your friends,” (S124)* being able to *“tell when they have read your message,” (S98), “using emojis to show my facial expression” (S100),* and *“voice recording features” (S38)* made it easier to talk about challenging topics. The ease of switching channels also helped facilitate difficult conversations, as *“there is no special features to follow to be on Facebook Messenger [from Facebook]” (S40),* and *“Messenger can be accessed on your phone without having to login to your Facebook account” (S29).*

However, even with the increased affordances and privacy of messaging platforms, participants felt that they could not replace face-to-face conversation in some instances, and they preferred to meet offline to resolve arguments that began online.

“What I have realized about using Internet to talk about some stuff is, usually you talk about it and then it doesn’t resolve right away, then my friends are like, ‘Well, let’s talk about it in person. It’s not going anywhere.’... When you talk about it in person, you can express your emotion[s] better, and they don’t read your emotion[s] wrong because you

Fights per Platform Compared to Hours Spent

Left axis displays percentage of users who have fought on the platform, right displays hours

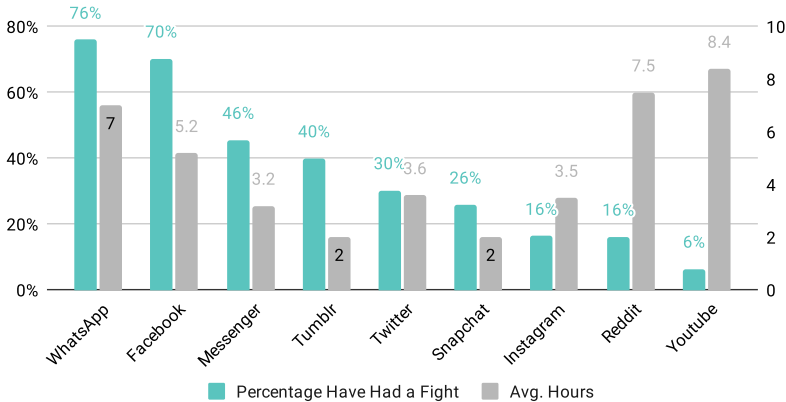


Fig. 2. Number of participants who have reported arguing on various social media platforms, compared to average hours spent per platform.

don't show your facial expression and stuff like that. So they talk about some stuff offline even though we started from KakaoTalk first" (P1).

Participants carefully select which platform to have a difficult conversation on, and they are open to channel switching to incorporate the affordances they most highly value during an argument (see Fig. 2 for an overview of where arguments occur). Their desire to channel switch also shows that the affordances which help when initiating an argument may not help as time progresses.

Quantifying Experiences in Public and Private Online Spaces. We used our survey to quantify participants' experiences arguing in public and private spaces. Using linear mixed models with participant ID and whether the platform was public or private as predictors, we analyzed agreement on a Likert scale from 1 (strongly disagree) to 5 (strongly agree) with the four statements in Table 2.

We found that participants were significantly more likely to agree with the statement, *There are things that I wish I could talk about on [platform] but don't because it might lead to an argument*, on public platforms ($M = 3.00$) compared to private platforms ($M = 2.72$). Participants were more likely to agree with the statement, *"I feel like I can talk about challenging or controversial topics on [platform]"*, private platforms ($M = 3.54$), rather than public platforms ($M = 3.12$). Thus, despite the fact that participants report becoming embroiled in arguments on both public and private platforms, they describe being more open and authentic in private spaces and more avoidant in public ones.

4.5 Image-Forward Platforms Have Fewer Arguments

Participants consistently reported that image-forward platforms such as Instagram, Snapchat, and Youtube led to fewer arguments than text-forward platforms. Participants explained, *"my Instagram conversations are more lighthearted"* (P10) because the purpose of interactions on the platform is to, *"send [other users] pictures"* (P10). Similarly, participants said, *"with Instagram, it's heavily photo-driven, so... it's not as easy to get into an argument"* (P16).

However, other participants believed that incorporating video and photo content could help to facilitate difficult conversations online. YouTube videos were highly regarded for narrowing the scope of the argument: *"I like the fact that one particular video can be focused on, without having to defend a whole issue"* (S48). YouTube videos also facilitated social cues since you can *"see the person's*

| Question | Public Mean (sd) | Private Mean (sd) | β | $F(df)$ | p | Cohen's d |
|--|---------------------|----------------------|--------------|---------------------------------------|------------------|-------------|
| <i>When I use [platform] I often find myself arguing with someone</i> | 2.19 (1.07) | 2.21 (1.16) | -0.019 | $F(1, 538) = 0.20$ | 0.65 | 0.02 |
| <i>I don't mind arguing with people on [platform]</i> | 2.73 (1.25) | 2.84 (1.21) | 0.036 | $F(1, 533) = 0.65$ | 0.42 | 0.09 |
| <i>There are things that I wish I could talk about on [platform] but don't because it might lead to an argument</i> | 3.00 (1.24) | 2.72 (1.26) | -0.18 | $F(1, 550) = 11.60$ | <0.001 | 0.20 |
| <i>I feel like I can talk about challenging or controversial topics on [platform]</i> | 3.12 (1.28) | 3.54 (1.20) | 0.17 | $F(1, 542) = 12.11$ | <.001 | 0.33 |

Table 2. Statistical analyses on survey responses regarding public and private social media platforms. Mean, standard deviation, and Cohen's d are captured from t-tests, β values are from the linear mixed model, and F and p values are from an ANOVA on the linear mixed model.

facial expressions on the videos they produce” (S56). The length of content was also considered beneficial, as *“you can make long videos to really explain yourself”* (S119).

Participants specifically mentioned how the visibility of people’s comments and text posts affected the frequency of arguments:

“With Instagram, it’s more like the picture, the person’s caption...and a lot of people don’t take the time to click on [the comments] and scroll through. Whereas, if you’re scrolling through Twitter, you just see it right there, and then you’re like, ‘Oh I don’t agree with that,’ and then people start arguing with each other... Twitter is designed more for arguments, just because when you see the first tweet that someone tweeted... that can spark an argument because it’s right there, they don’t have to click on like comments and see what someone else said.” (P17)

Survey participants corroborated interviewees’ reports of having fewer arguments on image-forward platforms. Across all participants, 7% said their last fight occurred on an image-forward platform and 93% said their last fight occurred on a text-forward platform, per categorization in Table 1. Participants consistently described the visibility of comments and conversation from other users as a feature that contributes to whether they view the platform as a place for arguments. Thus, manipulating the visibility of comments and conversations is likely an effective mechanism for foregrounding or foreclosing the possibility of discussing controversial topics.

4.6 Post and Comment Organization Facilitates Conversational Flow

Survey participants reported that the organization of a comments section can either help facilitate or hinder having difficult conversations online. For instance, Messenger, WeChat, and WhatsApp allow for *“one on one conversations, [which] keeps the focus on the discussion at hand”* (S60). This is in direct contrast to public social media such as Twitter, in which *“it’s easy to get lost in the noise”* (S111) and *“it’s hard to keep track of many conversations”* (S106). Participant’s often blamed this on the interface and organization of such platforms. *“Reddit’s reputation system makes it so that only the most popular topics are visible while controversial topics are quashed or simply not shown at all”* (S23). Regarding Facebook, *“One might be in the last chain of replies of a single comment, so it gets truncated automatically”* (S19). Or simply, in Twitter’s case, *“scrolling through a conversation feed can be annoying”* (S112). YouTube and Instagram received similar criticism on the lack of organization for discussion. *“Because of the way [YouTube] is set up, it feels more like leaving a review than having a conversation,”* (S94) and *“if you really want to discuss a topic and not just leave your opinion, the lack of interaction is an issue [on YouTube]”* (S17). Put simply, users *“just post and never respond [on Instagram]”* (S18). Thus, when a platform’s focus is video and photo content, some users’ desires to discuss content are hindered by the lack of interaction in the comments.

4.7 Users Value Less Moderation and Censorship from a Platform

Many participants described how low levels of moderation facilitated difficult discussions online across many platforms. People talked about the ability to *“comment freely without persecution”* (S75) and *“say what I want”* (S75) on YouTube, if moderation wasn’t enforced. They further explained that, *“if the [YouTube] channel doesn’t moderate, it’s pretty nice to talk in the live chat”* (S78). On Instagram, they said, *“everyone can post anything they want without restriction from moderators”* (S108). Users said that Reddit allowed them to have an *“open voice”* where *“anyone can share anything,”* (S123) and *“you can go back and forth as long as you want without limitations to the number of replies”* (S119).

Moderation by a platform was generally perceived as hindering difficult discussions. For instance, *“YouTube removes videos at their own discretion, and it’s usually a ridiculous or nonsensical reason”* (S113). It was also a hindrance that *“a lot of controversial videos have the comments section turned*

off” (S60), and “[YouTube] creators can moderate all comments, and they just filter things they don’t like” (S78). Twitter and Reddit’s moderation were perceived to be too heavy-handed, and it caused mistrust in how moderation and censorship decisions were made. Users said that they “don’t like censorship beyond blocking curse words” (S78), and “it’s very easy to fall into the ‘hate speech’ trap [on Twitter]” (S17). Users experienced frustration and mistrust when their content was censored. “You don’t know when someone at Twitter decides they don’t like my point of view or opinion and decides it’s hate speech...” (S113). Similarly, “Reddit moderators tend to ban users or close subreddits that the organization doesn’t agree with” (S104). On Messenger, participants said it was a hindrance that “they will time you out if you talk too much,” (S78) but “[the platform] has good options to block [other users]” (S3). People also said it was a benefit to themselves to “block people that are rude [on Instagram]” (S137), “turn off anonymous comments [on Tumblr]” (S96), and “censor groups and my posts using the privacy tool [on Instagram]” (S36).

Overall, users thought that the freedom to post their views on a variety of topics helped facilitate difficult conversations online. However, moderation, and especially moderation that is unexplained or viewed as biased, hurts open dialogue. Users appreciate being able to curate their own interactions with other by using a block or unfollow feature but do not appreciate being censored by a platform.

5 RESULTS: NOVEL DESIGNS AND EVALUATIONS

In interviews, participants brainstormed design ideas to support users while in the midst of a difficult discussion online. We clustered these ideas into 12 different design approaches (see Table 3). The design ideas were generated and analyzed separately from any particular platform. We used a repeated-measures ANOVA to measure the reactions (from very negative to very positive) of an additional 98 participants to the storyboards (see Figure 3). This revealed a highly significant difference in reaction to the 12 concepts ($F(1) = 72.45, p < .001, \eta^2 = .43$). This was mirrored in participants’ responses to the prompt, “I would be interested in trying [design]” ($F(1) = 27.24, p < .001, \eta^2 = .26$). This indicates that, while controlling for individual differences, certain storyboards were viewed more positively than others. To evaluate which design concepts might be of interest to users, we ran one-sample t -tests on the responses to each storyboard, evaluating each relative to a hypothesized population mean of three (the neutral score on our five-point scale). Six storyboards had an average score that was significantly higher than this neutral score and represented the categories of: *deleting content*, *blocking users*, *democracy*, *humanizing*, *channel switching*, and *emojis*. Several of these storyboards represented categories that are present in existing platforms, which may have influenced participants’ reactions to them.

5.1 What Users Fear in Interventions to Support Constructive Arguments

We identified four themes from participants’ feedback regarding the challenges of supporting constructive conversation through design. These concerns arose in response to many different interventions, suggesting they are inherent to the challenge of designing for online conflict, rather than any particular approach.

- *Dilution of intent*. Participants expressed concern that an intervention would dilute or alter their intended message. They explained that they did not want to use interventions that require sacrificing some of the meaning in their own messages or the messages of others because “it means the conversation is not true and honest.” Participants were concerned about the storyboard for *reasoning*, saying, “this is terrible in that it takes what the person is trying to say and essentially turns it into a form letter which no one likes.” They pushed back on an intervention to enforce reading a message before responding to it (*rules of engagement*), saying, “if you are trying to make a strong point, and the app slows you down, you might

Taxonomy of Design Approaches

| |
|--|
| Deleting Content. Deleting posts, comments, or other content that the user has created. |
| Blocking Users. Temporarily or permanently blocking a user or severing a connection. |
| Democracy. Leveraging community reactions to elevate content that promotes constructive dialogue, for example, through upvoting and downvoting, allowing users to flag posts as particularly constructive, or making algorithmic decisions in response to crowd input. |
| Humanizing. Building more empathy between parties by providing details about users, such as identity, background, current status, or mood, increasing the size of their profile picture, and preventing users from remaining anonymous. |
| Channel Switching. Providing the ability to move conversations from public to private channels or from text-based to voice or video calling. These were the most common interventions participants mentioned. |
| Emoticons. Adding emotional signals to the delivery of a user’s message by adjusting their use of emoticons. These interventions included approaches like removing angry emoticons, prompting users to use emoticons, or suggesting emoticons based on the text of a user’s message. |
| Censorship. Removing or replacing inflammatory content by suggesting word replacements, warning users with excessive word counts to prevent long-winded tirades, filtering particular words or parts of posts, and algorithms to filter out racial slurs. |
| Speed Bumps. Slowing down arguments in the moment, for example, by introducing a waiting period before a comment is posted or by enforcing fully reading a message before replying. These interventions sought to increase shared understanding between communication partners by nudging them to pay careful to their own words and the words of others. |
| Reflection. Detecting arguments and resurfacing them later to address regret users might feel, ask about the status of a relationship, or otherwise prompt them to work to strengthen or repair existing relationships. Users postulated that these designs might help users reconcile with friends once they have had time and emotional distance from an argument. |
| Reasoning. Providing users with facts that might encourage more shared understanding, such as tools for including data, connecting users to factual sources of information, and nudging users away from emotional argumentation and toward factual argumentation. |
| Rules of Engagement. Imposing restrictions on the way users interact, for example, by disabling caps lock, allowing only one person to type at a time, and enforcing turn-taking. Unlike censorship solutions, where messages were altered after they were submitted by the user, these interventions placed restrictions on the user as they composed their message. |
| Biofeedback. Leveraging biological feedback to better understand participants’ experiences during arguments. By detecting their heart rate or skin responses, participants’ interventions sought to provide just-in-time nudges to keep arguments constructive and remind participants of the effects of their words. |

Table 3. Design ideas generated in interviews distilled into a taxonomy, listed from most to least positively evaluated.

Reactions and Willingness to Try Storyboard Interventions

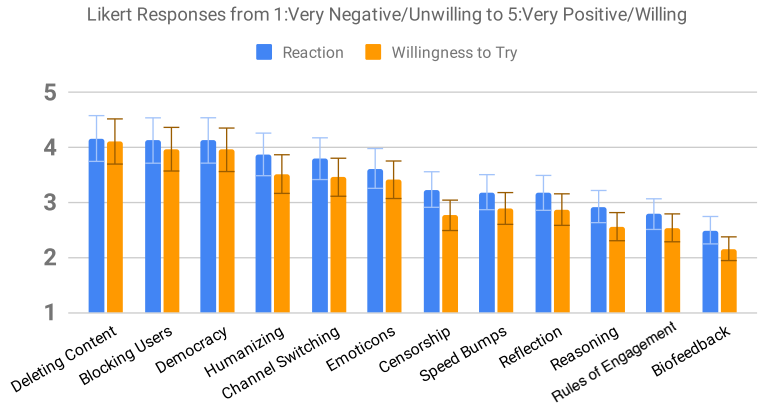


Fig. 3. Reactions and willingness to try to storyboard interventions generated from taxonomy.
Proc. ACM Hum.-Comput. Interact., Vol. 5, No. CSCW1, Article 156. Publication date: April 2021.

not be able to make your point as you intended." In these cases, participants explained that the purpose of these platforms is to allow users to authentically express themselves to one another, and impeding their ability to do so defeats the inherent purpose of the experience.

- *Backfire.* In response to many storyboards, participants expressed concern that the intervention intended to help might actually inflame the situation. Participants said that automation to resurface arguments and prompt reconciliation (*reflection*) might *"be adding salt to injury,"* and a template to help structure dialogue (*rules of engagement*) would lead people to *"get more pissed off."* They worried that *speed bumps* prompting users to pause and think would escalate the situation, saying that *"the person already in tension will become more[so] if the message is not sent immediately."* They also worried that a feature to *block* other users might make conflict worse, saying, *"blocking in the heat of the moment can cause more friction than necessary between friends."* And they worried about interventions that leverage *emoticons* saying, *"if anything, it might make things worse by implying a lack of seriousness in the topic."*
- *Intrusion.* Participants pushed back on a variety of interventions because of the intrusive data collection that would be required. They explained that even designs they found promising seemed problematically invasive, for example, *"I don't think I would like to report my own mood. I think it's very intrusive as well and would rather people not know,"* and, *"I still don't like that what you type is being monitored."* They raised concerns about using *biofeedback* to provide just-in-time support, saying, *"I think it is intrusive and creepy to detect someone's heart rate."*
- *Overkill.* Finally, participants disliked many interventions for their heavy-handedness, explaining that such interventions were not necessary to ensure strong interpersonal relationships, and the user should *"just say what they're thinking."* In these cases, participants described management from technology as, *"just unnecessary and too much."*

5.2 What Users Value in Interventions to Support Constructive Arguments

Despite these concerns, participants also appreciated many of the interventions, as demonstrated by their active interest in trying half of them. We examined their open-ended responses for themes in what they found valuable in these designs:

- *Privacy.* Many participants said that they valued support for easy *channel switching* and moving conversations into a private space when they became heated, saying, *"I like that it intervenes to suggest taking the conversation elsewhere."* Although this theme only came up in response to channel switching, we chose to report it here because an overwhelming majority of participants demonstrated how much they would value having this support. Participants described many benefits of moving to a private space, explaining, *"this way, people don't get annoyed and included in ongoing discussion that doesn't really involve them,"* and *"this would save a lot of people embarrassment from arguing in public."*
- *Empowerment.* Across interventions, participants said they appreciated when a design empowered them to customize their interactions and shape their communities, even as differences in opinion might persist. In response to *censorship*, which allowed users to filter out harsh language, one participant explained, *"I like the idea a lot. I'm offended by cursing and it causes me to lose sight of the message that they are trying to send."* Other participants said that using *democracy* to mark others' comments and posts as constructive would be valuable *"because sometimes really good, constructive comments can get lost in the chaos."* Across interventions,

participants valued nudges that helped the conversation retain respectful properties, appreciating when “*good messages gain visibility*” and expressed interest in designs that “*will help keep the conversation more civil*” by allowing them to take charge of the interaction.

- *Real Life Inspiration.* Participants appreciated interventions that drew inspiration from offline interactions and attempted to mimic aspects of productive in-person argumentation. Participants told us they liked seeing other users’ moods, saying, “*I like that it restores some of the social cues that we rely on in face-to-face communication, and that we miss in online chats,*” and, “*it makes the chat more like real life where you interact with people’s facial expressions, not just their words.*” Participants appreciated a mechanism to enforce reading messages before responding by comparing it to real life processes: “*I like that it is like active listening in a way.*” And they appreciated the ability to delete content, saying, “*I like that it’s an easy way to simply stop talking about an issue, like walking away in person.*” In these and other instances, participants drew upon what they appreciate about face-to-face arguments and explained that they were interested in these attributes being translated to a digital environment.
- *Common Understanding.* Across interventions, participants told us they appreciated when the design helped users reach more common ground and appreciate each other’s perspective. They liked when an intervention “*helps someone realize how their communication is being used,*” and they were interested in trying a feature that resurfaces old arguments (*reflection*) because “*it makes sense to be able to go over the argument and see what you were even fighting about in the first place.*” Many participants felt that emoticons would help participants cultivate a more precise understanding of each other’s emotional state, saying things like, “*I think sometimes things can be misconstrued through chat, and emoticons are helpful in letting people know what the intent behind the message is.*”

6 DISCUSSION

The majority of our participants said that they regularly use platforms where they want to discuss topics they currently avoid—ranging from politics to details about their personal life. They expressed the belief that working through disagreements can ultimately strengthen their ties and, as a result, said they want to engage in these hard discussions. And yet, many described avoiding any conversation with even a hint of controversy. Here, we discuss strategies to successfully take on the challenge of designing to support difficult conversations online, and we point to a need for an *interpersonal design* approach that foregrounds users’ relationships in the design process.

6.1 Showcasing the Need for an Interpersonal Design Approach

Designing to support constructive arguments is both a difficult and worthy design problem, as shown by participants’ unmet desire to have hard conversations online, their frustration with current platforms, and their wariness about many of the possible interventions we showed them. They surfaced concerns that reflected the precariousness of carrying out an argument well: they worried that one misstep from an over-eager intervention would backfire and make an argument worse than ever, and they worried that any change to the meaning of their message would erode the shared understanding that people must work so hard to build when arguing. Thus, although they see a need for designs that support this scenario, they also make clear that *no* intervention is better than an intervention that backfires, illustrating the difficulty of engaging with this design problem. These concerns are reflective of the fact that it is not only the argument that is at stake in these moments, but the relationship. Further, for our participants, the relationship often greatly outweighs the argument in its importance. Their choice to avoid arguments with friends and family online is often a sacrifice they make to protect their relationships with the people they love.

Numerous studies document the outsized influence interpersonal relationships have on long-term well-being [23, 25], concluding that close relationships are perhaps “*the greatest single cause*” of happiness [3]. Participants’ dissatisfaction with the status quo suggests that the design of the platforms they currently use, and the interventions we proposed, do not sufficiently prioritize their relationships. Designers have an opportunity to better account for this central part of life by foregrounding interpersonal relationships during the design process and drawing on evidence-based techniques for supporting healthy engagement in interpersonal interactions, an approach we refer to as *interpersonal design*. Interpersonal design represents a paradigm shift from supporting users’ goals for interacting with technology, to instead, supporting users’ goals in relating *to one another* through technology. Future works could focus on developing methodology that would support an interpersonal design approach, such as new research methods that incorporate pairs of participants and how relationship science can inform the design process.

6.2 Designs for Constructive Arguments

Our work shows that even though people do not often change their opinions [51], the process of talking through difficult topics is highly valued. Therefore, we encourage designers to judge an argument’s success by the level of authentic, good-faith, and respectful engagement users show. Below we discuss several mechanisms for encouraging prosocial engagement during online conflict: empowering communities, supporting vulnerability, and appropriate timing of interventions.

6.2.1 Empowering Communities with Collective Online Action. Users’ desire to drive the visibility of content was reflected in their perception of moderation, censorship, and content organization on current social media platforms, which they considered frustrating impediments to discussing challenging topics. Prior works have also found that heavily moderating communities can lead to self-censorship [56] and exclude underrepresented populations from the deliberation process [54]. Therefore, interventions that enforce civility unilaterally or optimize for popular content, as many platforms do, may do so by sacrificing equity. This makes sense, as changes to existing power structures and the status quo will necessarily upset some people and lead to disagreement. Designers can work toward empowering users and promoting successful discussion of challenging topics by including users in the algorithmic design and moderation decisions, in addition to designing equitable top-down moderation (e.g. [13]).

Including users in algorithmic design and moderation could include more heavily weighting user feedback through *democracy*, one of our most highly rated storyboards. *Democracy* leverages community input to: 1) boost comments that are particularly constructive and 2) reduce the visibility of more inflammatory comments. Designers can similarly create interfaces that offer lightweight ways for users to contribute to collective moderation. For example, Brewer et al. [9] launched a community-centered moderation effort to combat harassment on Twitch, and found overwhelming interest and support, corroborating our participants’ desires for communal norms-setting in online communities. They found that the moderation campaign cultivated meaningful change in user behavior and lead to more constructive dialogue regarding community norms. Other future directions might provide a feature that enables bystanders to participate constructively or facilitate easy ways of rallying around constructive content, similar to Taylor et al. [52] and Kriplean et al. [29]. Users clearly value the ability to shape their communities and content consumption, but they want to be empowered to make these decisions for themselves, rather than have their intentions diluted or otherwise mismanaged by technology.

6.2.2 Interface Support for Vulnerability. When users are performing an action in which personal responsibility for the outcome is important or when the stakes are high, technology should generally seek to *augment* experiences rather than *automate*, if it intervenes at all [16]. Our research shows

that online arguments can be high-stakes interpersonal interactions, and people prefer to remain in control and responsible during these delicate exchanges. Therefore, designers choosing to intervene should proceed with caution. At the same time, our results show that social media designers and researchers should not refrain from taking action either, as users were excited about the possibility for technology to help build more common understanding and to support their relationships.

Two of the most highly valued design approaches were *channel switching* to a private space and *humanizing* others. *Channel switching* replicates a natural tendency participants already have to resolve disagreements privately, especially if they began in a group discussion or on a public platform. A salient feature enabling users to seamlessly move a discussion to a private messaging channel may help prevent unnecessarily damaging “comment wars” on public platforms [15], and make it easier for users to save face, be authentic, and express sentiments that might make them feel vulnerable. Participants felt that *humanizing* other users by adding more details about their identity, mood, and background would help to prevent online conflicts from becoming toxic. These interventions may be more successful as persistent, ambient features that continually broadcast humanizing information, rather than just-in-time interventions that appear when an argument becomes heated. Because although it is possible to detect arguments from text [41, 50], doing so in real-time carries a risk participants identified of appearing “*intrusive and creepy*.”

When assessing users’ perceptions of the risks and benefits of design interventions, we found that users wanted certain information from others that they were not comfortable providing themselves. In fact, the highly rated *humanizing* and lowest rated *biofeedback* interventions displayed similar data to develop insight into others’ moods, however, there was one critical difference. *Humanizing* displayed emotional data without specifying how the data would be gathered, whereas *biofeedback* illustrated how a user would enter their heart rate into the app. This suggests that users desire data from others that they paradoxically are not comfortable providing. In face-to-face contexts, people fluidly express or repress this non-verbal, emotional information through facial expressions, tone, and other subconscious gestures [14, 17, 40]. Turning non-verbal, emotional cues into explicitly shared data requires an additional level of conscious thought, trust, and acceptance of emotions that users may be disincentivized or even ill-equipped to provide if they are working to repress an uncomfortable emotion. (For instance, it may be easier to subtly roll your eyes and sigh than to type that you are feeling exasperated into a mood prompt.) Future work is needed to explore how users can incorporate emotional cues into their social media use, without intruding or inappropriately broadcasting users’ sensitive emotional information.

6.3 Considering the Impact of Stages of an Argument

When initiating difficult discussions, participants most valued affordances that helped them remain anonymous or made their messages ephemeral. However, at later stages in a difficult discussion, and especially when working to find a resolution, these features became less valuable, and users instead wanted features that surface social cues and support one-to-one conversations through messaging, voice and video calling, and even encouraging offline interactions. This has important implications for research, as it indicates user feedback on communication platforms is highly dependent on the context and stage of an argument in which the feature is used. For instance, users discussed the value of resurfacing old arguments through *reflection*, but also that it could make a past negative outcome feel worse. This suggests that *reflection* may be valuable to users after sufficient time and emotional distance, but it could backfire if it occurred too soon. Therefore, we suggest that future research on online arguments differentiate feedback based on phase of argumentation, such as *initiating* compared to *resolving*.

A limitation of this work is that we do not know what drives participants’ perceptions of different affordances, both across platforms and phase of conflict. It is possible that universally, participants

prefer to incorporate more affordances over the course of an argument. However, it is also possible that some users will generally prefer more anonymous and ephemeral online spaces, whereas others prefer to incorporate more affordances over the course of an argument that include emotional cues one would receive in-person. For instance, it is possible that those who prefer more anonymity may be more avoidant or accommodating in conflict, whereas those who desire more emotional feedback may be more collaborative in their approach to conflict management. Avoidant styles in conflict are characterized by a lack of direct communication, including hinting, joking, and passive aggression. Accommodating is also passive, in which someone complies with another without providing their own input. The collaborative style involves a high degree of concern for self and others, and usually indicates investment in the conflict and relationship. Those with a collaborative style are more likely to remain flexible in conflict, carefully listen and provide verbal and non-verbal feedback, and identify areas of common ground [22, 43]. These behaviors are promoted in the highly rated *humanizing* intervention, and across interventions, users were excited about the potential for technology to build common understanding. Future works could further explore how technology could promote collaborative styles of conflict management.

7 CONCLUSION

This work contributes a mixed-methods investigation of users' experiences discussing difficult topics across popular platforms and an analysis of novel user-generated designs to support relationships during online conflict. We show that there is an unmet desire to discuss challenging topics in online spaces and how various affordances of popular platforms help and hinder such discussions. We find that participants were most appreciative of designs that empower them to shape their own interactions and those of their community, and they simultaneously fear interventions that could dilute their intended message, backfire by making arguments worse, or unnecessarily intrude. Users are excited by the possibility of technology helping them reach more common ground during online conflicts, while also showcasing the need for caution when intervening in the midst of conflict. Our findings highlight a need for design to focus on *relationships* in addition to user experiences, an approach we label *interpersonal design*. By creating effective methods for systematically centering users' relationships in the design process, the research community can arm designers with a powerful tool for caring for their users and, in doing so, support the hard but necessary work of arguing well.

ACKNOWLEDGMENTS

This work was funded in part by a gift from Facebook to Alexis Hiniker. We would like to thank [xkcd comics](#) for the inspiration for our title. We thank our reviewers for their advice, feedback, and support, which greatly strengthened our paper. Thank you to Samuel Ainsworth, who read multiple versions of this paper and provided excellent advice.

REFERENCES

- [1] 2019. Social Media Fact Sheet. <https://www.pewresearch.org/internet/fact-sheet/social-media/>
- [2] Henk Aarts and A. P. Dijksterhuis. 2003. The silence of the library: environment, situational norm, and social behavior. *Journal of personality and social psychology* 84 1 (2003), 18–28.
- [3] Michael Argyle. 2013. *The psychology of happiness*. Routledge.
- [4] Mahmoudreza Babaei, Juhi Kulshrestha, Abhijnan Chakraborty, Fabricio Benevenuto, Krishna P. Gummadi, and Adrian Weller. 2018. Purple Feed: Identifying High Consensus News Posts on Social Media. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (New Orleans, LA, USA) (AIES '18). Association for Computing Machinery, New York, NY, USA, 10–16. <https://doi.org/10.1145/3278721.3278761>
- [5] Liad Bareket-Bojmel, Simone Moran, and Golan Shahar. 2016. Strategic self-presentation on Facebook: Personal motives and audience response to online behavior. *Computers in Human Behavior* 55 (2016), 788 – 795. <https://doi.org/10.1016/j.chb.2015.10.033>
- [6] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and Its Consequences for Online Harassment: Design Insights from HeartMob. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 24 (Dec. 2017), 19 pages. <https://doi.org/10.1145/3134659>
- [7] Lindsay Blackwell, Nicole Ellison, Natasha Elliott-Deflo, and Raz Schwartz. 2019. Harassment in Social Virtual Reality: Challenges for Platform Governance. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 100 (Nov. 2019), 25 pages. <https://doi.org/10.1145/3359202>
- [8] Porismita Borah. 2014. Does it matter where you read the news story? Interaction of incivility and news frames in the political blogosphere. *Communication Research* 41, 6 (2014), 809–827.
- [9] Johanna Brewer, Morgan Romine, and TL Taylor. 2020. Inclusion at Scale: Deploying a Community-Driven Moderation Intervention on Twitch. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 757–769.
- [10] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) (CSCW '17). Association for Computing Machinery, New York, NY, USA, 1217–1230. <https://doi.org/10.1145/2998181.2998213>
- [11] Danielle Keats Citron. 2014. *Hate crimes in cyberspace*. Harvard University Press.
- [12] John W Creswell and Cheryl N Poth. 2016. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.
- [13] Elizabeth Dwoskin, Natasha Tiku, and Heather Kelly. 2020. Facebook is overhauling its hate speech algorithms. <https://www.washingtonpost.com/technology/2020/12/03/facebook-hate-speech/>. (Accessed on 02/15/2021).
- [14] Paul Ekman. 2009. Lie catching and microexpressions. *The philosophy of deception* 1, 2 (2009), 5.
- [15] Jesse Fox and Jennifer J. Moreland. 2015. The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances. *Computers in Human Behavior* 45 (2015), 168 – 176. <https://doi.org/10.1016/j.chb.2014.11.083>
- [16] Google. [n.d.]. People + AI Guidebook. <https://pair.withgoogle.com/chapter/user-needs/>
- [17] John Mordechai Gottman and Robert Wayne Levenson. 2002. A two-factor model for predicting when a couple will divorce: Exploratory analyses using 14-year longitudinal data. *Family process* 41, 1 (2002), 83–96.
- [18] Catherine Grevet, Loren G. Terveen, and Eric Gilbert. 2014. Managing Political Differences in Social Media. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) (CSCW '14). Association for Computing Machinery, New York, NY, USA, 1400–1408. <https://doi.org/10.1145/2531602.2531676>
- [19] Carla F. Griggio, Joanna McGrenere, and Wendy E. Mackay. 2019. Customizations and Expression Breakdowns in Ecosystems of Communication Apps. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 26 (Nov. 2019), 26 pages. <https://doi.org/10.1145/3359128>
- [20] Barna Group. 2017. *Why People Fight Online*. <https://www.barna.com/research/people-fight-online/>
- [21] Jürgen Habermas. 1984. *The theory of communicative action: Reason and the rationalization of society*. Heneimann.
- [22] Owen Hargie. 2010. *Skilled interpersonal communication: Research, theory and practice*. Routledge.
- [23] Linda M Hartling. 2008. Strengthening resilience in a risky world: It's all about relationships. *Women & Therapy* 31, 2-4 (2008), 51–70.
- [24] Sameer Hinduja and Justin W. Patchin. 2008. Cyberbullying: An Exploratory Analysis of Factors Related to Offending and Victimization. *Deviant Behavior* 29, 2, 129–156. <https://doi.org/10.1080/01639620701457816>
- [25] Julianne Holt-Lunstad, Timothy B Smith, and J Bradley Layton. 2010. Social relationships and mortality risk: a meta-analytic review. *PLoS medicine* 7, 7 (2010), e1000316.
- [26] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2019. Online Harassment and Content Moderation: The Case of Blocklists. *ACM Trans. Comput.-Hum. Interact.* 25, 2, Article 12 (2019). <https://doi.org/10.1145/3185593>

- [27] Shagun Jhaver, Pranil Vora, and Amy Bruckman. 2017. *Designing for Civil Conversations: Lessons Learned from ChangeMyView*. Technical Report. Georgia Institute of Technology.
- [28] Nicholas A. John and Shira Dvir-Gvirsman. 2015. "I Don't Like You Any More": Facebook Unfriending by Israelis During the Israel–Gaza Conflict of 2014. *Journal of Communication* 65, 6 (2015), 953–974. <https://doi.org/10.1111/jcom.12188> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcom.12188>
- [29] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012. Supporting Reflective Public Thought with Considerit. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) (CSCW '12). Association for Computing Machinery, New York, NY, USA, 265–274. <https://doi.org/10.1145/2145204.2145249>
- [30] Travis Kriplean, Michael Toomim, Jonathan Morgan, Alan Borning, and Andrew Ko. 2012. Is This What You Meant? Promoting Listening on the Web with Reflect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI '12). Association for Computing Machinery, New York, NY, USA, 1559–1568. <https://doi.org/10.1145/2207676.2208621>
- [31] Nicole Krämer, Laura Hoffmann, and Sabrina Eimler. 2015. Not Breaking Bonds on Facebook–Mixed–Methods Research on the Influence of Individuals' Need to Belong on 'Unfriending' Behavior on Facebook. *International Journal of Developmental Science* 9 (08 2015), 61–74. <https://doi.org/10.3233/DEV-150161>
- [32] Haewoon Kwak, Jeremy Blackburn, and Seungyeop Han. 2015. Exploring cyberbullying and other toxic behavior in team competition online games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3739–3748.
- [33] 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 (2012), 434–443.
- [34] Q. Vera Liao and Wai-Tat Fu. 2013. Beyond the Filter Bubble: Interactive Effects of Perceived Threat and Topic Involvement on Selective Exposure to Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 2359–2368. <https://doi.org/10.1145/2470654.2481326>
- [35] Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and J. Brown. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology* 25, 2 (2006), 115–126. <https://doi.org/10.1080/01449290500330448>
- [36] Maciek Lipinski-Harten and Romin W. Tafarodi. 2013. Attitude moderation: A comparison of online chat and face-to-face conversation. *Computers in Human Behavior* 29 (2013), 2490–2493.
- [37] Zhe Liu and Ingmar Weber. 2014. *Is Twitter a Public Sphere for Online Conflicts? A Cross-Ideological and Cross-Hierarchical Look*. Springer International Publishing, Cham, 336–347. https://doi.org/10.1007/978-3-319-13734-6_25
- [38] Mirca Madianou and Daniel Miller. 2013. Polymedia: Towards a new theory of digital media in interpersonal communication. *International Journal of Cultural Studies* 16, 2 (2013), 169–187. <https://doi.org/10.1177/1367877912452486>
- [39] Suman Kalyan Maity, Aishik Chakraborty, Pawan Goyal, and Animesh Mukherjee. 2018. Opinion Conflicts: An Effective Route to Detect Incivility in Twitter. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 117 (Nov. 2018), 27 pages. <https://doi.org/10.1145/3274386>
- [40] Albert Mehrabian et al. 1971. *Silent messages*. Vol. 8. Wadsworth Belmont, CA.
- [41] Amita Misra, Pranav Anand, Jean Fox Tree, and Marilyn Walker. 2017. Using Summarization to Discover Argument Facets in Online Ideological Dialog. (09 2017).
- [42] Parviz Molavi, Niloofar Mikaeili, Mohammad Ali Ghaseminejad, Zhila Kazemi, and Misagh Pourdonya. 2018. Social Anxiety and Benign and Toxic Online Self-Disclosures: An Investigation Into the Role of Rejection Sensitivity, Self-Regulation, and Internet Addiction in College Students. *The Journal of nervous and mental disease* 206, 8 (2018), 598–605.
- [43] John Oetzel, Adolfo J Garcia, and Stella Ting-Toomey. 2008. An analysis of the relationships among face concerns and facework behaviors in perceived conflict situations. *International Journal of Conflict Management* (2008).
- [44] Elia Powers, Michael Koliska, and Pallavi Guha. 2019. "Shouting Matches and Echo Chambers": Perceived Identity Threats and Political Self-Censorship on Social Media. *International Journal of Communication* 13 (2019), 20.
- [45] Bryan Semaan, Heather Faucett, Scott P. Robertson, Misa Maruyama, and Sara Douglas. 2015. Designing Political Deliberation Environments to Support Interactions in the Public Sphere. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 3167–3176. <https://doi.org/10.1145/2702123.2702403>
- [46] Bryan C. Semaan, Scott P. Robertson, Sara Douglas, and Misa Maruyama. 2014. Social Media Supporting Political Deliberation across Multiple Public Spheres: Towards Depolarization. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) (CSCW '14). Association for Computing Machinery, New York, NY, USA, 1409–1421. <https://doi.org/10.1145/2531602.2531605>

- [47] Christopher Sibona. 2014. Unfriending on Facebook: Context Collapse and Unfriending Behaviors. In *Proceedings of the 2014 47th Hawaii International Conference on System Sciences (HICSS '14)*. IEEE Computer Society, USA, 1676–1685. <https://doi.org/10.1109/HICSS.2014.214>
- [48] Abhay Sukumaran, Stephanie Vezich, Melanie McHugh, and Clifford Nass. 2011. Normative Influences on Thoughtful Online Participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing Machinery, New York, NY, USA, 3401–3410. <https://doi.org/10.1145/1978942.1979450>
- [49] John Suler. 2004. The Online Disinhibition Effect. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* 7 3 (2004), 321–6.
- [50] Reid Swanson, Brian Ecker, and Marilyn Walker. 2015. Argument Mining: Extracting Arguments from Online Dialogue. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Prague, Czech Republic, 217–226. <https://doi.org/10.18653/v1/W15-4631>
- [51] 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* (Montréal, Québec, Canada) (*WWW '16*). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 613–624. <https://doi.org/10.1145/2872427.2883081>
- [52] Samuel Hardman Taylor, Dominic DiFranzo, Yoon Hyung Choi, Shruti Sannon, and Natalya N. Bazarova. 2019. Accountability and Empathy by Design: Encouraging Bystander Intervention to Cyberbullying on Social Media. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 118 (Nov. 2019), 26 pages. <https://doi.org/10.1145/3359220>
- [53] TEDx. 2016. The beauty of conflict | Clair Canfield | TEDxUSU. Retrieved from https://www.youtube.com/watch?v=55n9pH_A0O8&ab_channel=TEDxTalks.
- [54] Matthias Trénel. 2009. Facilitation and inclusive deliberation. (2009).
- [55] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. “I Regretted the Minute I Pressed Share”: A Qualitative Study of Regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania) (*SOUPS '11*). Association for Computing Machinery, New York, NY, USA, Article 10, 16 pages. <https://doi.org/10.1145/2078827.2078841>
- [56] Scott Wright. 2009. The Role of the Moderator: Problems and Possibilities’.
- [57] Sheng Wu, Tung-Ching Lin, and Jou-Fan Shih. 2017. Examining the antecedents of online disinhibition. *IT & People* 30 (2017), 189–209.
- [58] Sarita Yardi and Danah Boyd. 2010. Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science, Technology & Society* 30, 5 (2010), 316–327. <https://doi.org/10.1177/0270467610380011> arXiv:<https://doi.org/10.1177/0270467610380011>
- [59] Xuan Zhao, Cliff Lampe, and Nicole B. Ellison. 2016. The Social Media Ecology: User Perceptions, Strategies and Challenges. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 89–100. <https://doi.org/10.1145/2858036.2858333>