'Walking Into a Fire Hoping You Don't Catch': Strategies and Designs to Facilitate Cross-Partisan Online Discussions

ASHWIN RAJADESINGAN, University of Michigan, Ann Arbor, USA
CAROLYN DURAN, University of Michigan, Ann Arbor, USA
PAUL RESNICK, University of Michigan, Ann Arbor, USA
CEREN BUDAK, University of Michigan, Ann Arbor, USA

While cross-partisan conversations are central to a vibrant deliberative democracy, these conversations are hard to have, especially amidst unprecedented levels of partisan animosity we observe today. We report on a qualitative study of 17 US residents who engage with outpartisans on Reddit to understand what they look for in these interactions, and the strategies they adopt. We find that users have multiple, sometimes contradictory expectations of these conversations, ranging from deliberative discussions to entertainment and banter. In aiming to foster 'good' cross-partisan discussions, users make strategic choices on which subreddits to participate in, who to engage with and how to talk to outpartisans, often establishing common ground, complimenting, and remaining dispassionate in their interactions. Further, contrary to offline settings where knowing more about outpartisan interlocutors help manage disagreements, on Reddit, users look to actively learn as little as possible about them for fear that such information may bias their interactions. However, through design probes, we find that users are actually open to knowing certain kinds of information about their interlocutors, such as non-political subreddits that they both participate in, and to having that information made visible to their interlocutors. However, making other information visible, such as the other subreddits that they participate in or their past comments, though potentially humanizing, raises concerns around privacy and misuse of that information for personal attacks especially among women and minority groups. Finally, we identify important challenges and opportunities in designing to improve online cross-partisan interactions in today's hyper-polarized environment.

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

Additional Key Words and Phrases: political discussions, affective polarization, partisanship, social media

ACM Reference Format:
Ashwin Rajadesingan, Carolyn Duran, Paul Resnick, and Ceren Budak. 2021. 'Walking Into a Fire Hoping You Don't Catch': Strategies and Designs to Facilitate Cross-Partisan Online Discussions. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 393 (October 2021), 30 pages. https://doi.org/10.1145/3479537

1 INTRODUCTION

Casual political conversations through which individuals construct their identities, explore alternate perspectives and form considered opinions are key to a deliberative democracy [34]. Many of these political interactions take place in social media where users discuss politics among other topics with

Authors' addresses: Ashwin Rajadesingan, University of Michigan, Ann Arbor, USA, arajades@umich.edu; Carolyn Duran, University of Michigan, Ann Arbor, USA, cduran@umich.edu; Paul Resnick, University of Michigan, Ann Arbor, USA, presnick@umich.edu; Ceren Budak, University of Michigan, Ann Arbor, USA, cbudak@umich.edu.

2573-0142/2021/10-ART393 $15.00
https://doi.org/10.1145/3479537

friends, acquaintances and often, strangers. Although online political interactions are associated with some positive outcomes such as increased civic participation [69], they are often unpleasant experiences; about 70% of social media users report feeling stressed and frustrated when discussing politics with others on social media that they disagree with [3]. Worryingly, the tone of political discussions online tends to be angrier, less civil and less respectful than offline conversations [18].

A major factor contributing to hostility in both online and offline political discussions is the heightened levels of affective polarization that we observe today, a tendency of partisans to view opposing partisans negatively and copartisans positively [31]. Increasingly, rank-and-file Republicans and Democrats view each other as selfish, hypocritical and close-minded [30]. This increased outparty animosity is explained by Social Identity Theory which argues that by merely categorizing individuals into groups (here, Republicans and Democrats), group identities are activated, creating an 'us' versus 'them' group dynamic [76]. Importantly, unlike protected attributes such as race where group-related behaviors are moderated by strong social norms and laws against discrimination, no such norms temper partisan hostility [30]. Thus, platform designers must account for and mitigate the deleterious effects of partisan identity when building systems that facilitate cross-partisan discourse.

Most prior research on improving cross-partisan discourse has predominantly aimed at addressing partisan bias in information consumption to burst filter bubbles [50, 52, 54], with little emphasis on mitigating the role of partisan identity during interactions. In this work, we aim to reduce partisan prejudice by designing interfaces showcasing user information to promote cross-categorization and decategorization—two strategies adopted from social psychology research on inter-group conflict. Cross-categorization *increases awareness of cross-cutting identities with members of the outgroup* [7]. Decategorization *increases awareness of the distinctiveness of individual members of the ingroup and outgroup* [6]. We conduct a qualitative study using semi-structured interviews (i) to first understand the expectations, concerns and strategies of users who engage in cross-partisan interactions and (ii) to seek feedback on designs and evaluate the types of information that can facilitate better cross-partisan discussions. We focus our analysis on Reddit, a popular social networking discussion site which hosts hundreds of political discussion communities (subreddits).

Our interviews reveal complex, and at times contradicting, motivations for participation in online cross-partisan talk, where participants look for serious deliberation but also entertainment and banter in political discussions. Participants also highlighted varied concerns with engaging in cross-partisan discourse. As one participant succinctly put it, cross-partisan talk can sometimes feel like "walking into a fire hoping you don't catch", requiring refined strategies to increase the odds of having compelling discussions. However, our designs to decategorize and cross-categorize users produced mixed effects. While participants expressed strong support for the cross-categorization inspired "shared subreddit" component, they—especially women and minorities—expressed that the extra user information provided by other components, while potentially humanizing, increased scrutiny on their profiles and would likely be used to attack them or derail discussions. We discuss the implications of these findings and detail the design challenges and opportunities to improve online cross-partisan discourse.

2 RELATED WORK

2.1 Partisan identity in online deliberation

Normative theories of deliberation largely stem from Jürgen Habermas' conception of the public sphere, where citizens engage in rational-critical argumentation to form public opinion [24]. The presupposed conditions central to such argumentation such as inclusion, discursive equality, ideal role-taking (impartiality and reciprocity) and absence of coercive power have been conceptualized as

ideals of deliberation by deliberative theorists [4]. While these ideals aim to ensure that individuals are swayed only by the best of arguments, in practice, empirical research reveals how partisan identities play a consequential role in how people engage with outpartisans and their arguments [26]. Motivated to maintain their party's positive distinctiveness and advance group status, partisans engage in inparty favoritism and outparty animosity [28]. This results in increased partisan hostility which we review below.

Partisan hostility typically manifests in the form of incivility and abuse targeted at outparty supporters. Partisans are more willing to denigrate outpartisans while judging incivility expressed by outpartisans more strongly than incivility by copartisans [58]. Exposure to copartisan attacks on outpartisans encourages copy-cat attacks while attacks by outpartisans result in stronger retaliation [22]. Further, this partisan hostility is often favored, even in highly moderated online discussion spaces; studying the heavy moderated New York Times comments section, researchers observed that uncivil partisan comments received more "recommendations" (similar to upvotes) from users than comments that contained only uncivil language or only partisan language [49]. Worryingly, exposure to partisan ad hominem criticism in news comments, which are exceedingly common online, result in more prejudiced attitudes towards outpartisans further exacerbating affective polarization [72].

These group-motivated behaviors may even be exacerbated in online spaces. The Social Identity Theory of Deindividuation Effects (SIDE) posits that visual anonymity afforded by online platforms because of the lack of visible individuating information about group members increases the salience of group identities and adherence to group normative behavior [61]. Following self-categorization theory, in the presence of an accessible group identity, individuals become depersonalized and view themselves and others less as individuals having distinct personalities but instead as interchangeable group members [77]. Although commonly used to explain behavior in intragroup contexts, similar group dynamics have been observed in intergroup contexts as well. Through a series of experiments in intergroup online settings where participants were anonymous except for their group membership labels, Postmes et al. [56, 57] showed that the depersonalization predicted by the SIDE model increased the relative salience of group boundaries and led to stereotyped perception of the outgroup. An important condition for observing depersonalization is that group identity must be accessible and salient during intergroup interactions. In the context of online political discussions, prior research overwhelmingly points to the salience of partisan identity in these interactions [67]. Moreover, in many subreddits such as r/AskTrumpSupporters and r/AskALiberal, users, who already have very little individuating information about them (such as a profile picture, bio-sketch etc.) are also required to use a tag to identify themselves as a 'Trump Supporter' or a 'Progressive', setting up the ideal conditions for group-motivated behavior where a lack of individuating information is coupled with the presence of group cues.

2.2 Cross-categorization and decategorization to reduce partisan hostility

Given that categorization into partisan groups forms the basis for partisan hostility, we review two social psychology approaches aimed at changing individuals' level of categorization: cross-categorization [7] and decategorization [6]. These strategies rely on the fact that individuals have multiple social identities apart from their partisan identities which may be activated to affect interaction dynamics [27].

2.2.1 Cross-categorization. Cross-categorization aims to make individuals of a group aware that they share membership in another dimension with individuals of the outgroup [7]. Revealing overlapping or shared group memberships makes social categorization more complex and reduces bias by increasing awareness of multiple subgroups within the outgroup [14]. Further, by making

cross-cutting identities more salient, assimilation effects of the cross-cutting identity tend to offset the discriminatory nature of the partisan identity. Studying how other identities interact with partisan identity, Mason [45] observed "a cross-cutting calm", individuals with cross-cutting identities (for example, secular Republicans and evangelical Democrats) significantly reduced angry responses to party threats, exhibiting anger at even lower rates than weak partisans. Recently, testing the effects of shared non-political identities on partisan hostility, Levendusky experimentally found that individuals exhibited significantly higher warmth (by over 20%) towards outpartisans when they were identified as supporting the same football team compared to no team identification ([40], Chapter 3). Based on these findings, we design an interface that surfaces "shared subreddits", users' shared membership in other nonpolitical communities, during their interaction to reduce hostility stemming from partisan identity. By explicitly highlighting shared group membership, we alert the user to the presence of "calming" cross-cutting identities.

2.2.2 Decategorization. Decategorization is aimed at increasing the salience of intragroup variability by highlighting the distinctiveness of individual members [13]. By exposing individuals to information about multiple other group memberships of outgroup members, individuals are nudged to differentiate outgroup members from the outgroup stereotype. Thus, by providing a more complex view of each outgroup member, individuals can evaluate them based on their personal merit rather than their stereotypical group memberships [6]. In politics, research suggests that people consistently stereotype outpartisans as being politically engaged extreme ideologues when no other information is provided about them which exacerbates outpartisan hostility [17]; when outpartisans were instead described as talking politics rarely and being ideologically moderate (who in reality is the modal outpartisan), outpartisans were evaluated more positively [17, 35]. Similarly, participants evaluated outpartisans who were less interested in politics more positively in a hypothetical roommate selection experiment [68]. These findings suggest that providing information contextualizing the extent of users' political versus non-political attachments may help reduce partisan hostility. Thus, in addition to highlighting shared subreddits in our design, we also provide non-political individuating information about outpartisans in the form of "active subreddits", non-political subreddits that the interlocutor has recently participated in. By explicitly highlighting the other group memberships, we aim to decategorize the user as solely a member of their partisan group, instead we showcase the user as a distinctive individual with varied interests and identities, unrelated to their political leanings.

Another intervention closely related to this work is the intergroup contact hypothesis. The contact hypothesis suggests that interpersonal interactions between outgroup members under certain conditions: equal status, common goals, cooperative and institutional support will reduce intergroup prejudice [55]. However, as Wojcieszak and Warner [80] note, the intergroup contact hypothesis has not been extensively tested in the context of partisanship. While intergroup contact is central to this study, we aim to facilitate positive intergroup contact by reducing partisan bias, whereas studies testing the intergroup contact hypothesis examine the effects of intergroup contact on reducing partisan bias.

2.3 Designing for online deliberation

Researchers aiming to improve political deliberation have typically focused on two aspects: diversifying information consumption [50, 51, 53] and facilitating deliberative interactions [37, 38]. As this work primarily concerns the latter, we review in detail the innovative interface designs that facilitate quality deliberation while reducing hostility in discussions. Early work on online deliberation centered around highly structured interactions mapping information into facts, positions, arguments and relationships between them [71]. In practice, these formal systems erected

high barriers to usage as they required training to help users navigate complex predetermined interaction structures and argumentation schema [70]. Over the past decade, researchers have aimed to facilitate high-quality deliberation while reducing such impediments, focusing on design considerations that center active contribution, navigability, usability, quality content and adoption [75]. Kriplean et al. [37] introduced ConsiderIt, a system that facilitates reflection of others' perspectives by allowing users to form their own pro/con list on a particular topic by also including pro/con points contributed by others. Kriplean and colleagues [38] also built Reflect, a commenting system that makes active listening the normative behavior for users of the system by including a small listening text box along with the comment for users to succinctly summarize the original comment. Another system, OpinionSpace [19] maps users to points in a 2-D space based on their responses to five general value-based questions (answers to these questions map to either liberal or conservative leaning opinions), with the distance between the points representing the similarity between user answers to the question set. When a user clicks on a point, they can rate how much they agree and respect a comment posted by the user corresponding to the point. These systems all aim to make conversations more reflective. Alternately, finding that users often used multiple social media platforms, Semaan et al. built Poli [64, 65], an integrated political deliberation environment that aggregates multiple social media.

2.3.1 Managing hostility in online deliberation. Hostility stemming from interactions have been typically handled in two ways: (i) by structuring interactions to reduce direct contact (for example, ConsiderIt uses pro/con lists instead of facilitating back and forth interactions between users) and (ii) by removing or sanctioning problematic content, users or even entire communities [8, 32, 63]. More recently, researchers have aimed to design interfaces to proactively reduce hostility. Seering et al. [62] designed psychologically embedded CAPTCHAs to prime users (just before replying) to trigger positive emotions that increased positivity, analytical complexity and interpersonal connectedness even in cross-partisan situations. Grevet et al. [23] studying how weak ties manage political differences on Facebook, recommend another proactive approach; they suggest that "making common ground visible (i.e., highlighting past interactions and shared interests) during contentious discussions could alleviate in-the-moment tension." This lends further support to our design choice to highlight shared subreddits during interactions. Although Reddit users are unlikely to know each other unlike Facebook, we expect that showing shared non-political group memberships will likely still have an effect of alleviating tension. Somewhat paradoxically, many of Reddit's design choices (e.g. up/down voting mechanisms) and participation cultures (e.g. circlejerking[1]) which contribute to the insularity of the subreddits may actually strengthen the effects of shared memberships in these communities by increasing users' bonds with other community members [2].

2.3.2 Managing partisan identities in online deliberation. Despite the prominence of partisanship in political interactions, most systems or designs (barring a few notable exceptions such as ConsiderIt and OpinionSpace) do not specifically address the prevailing group dynamics in these interactions. ConsiderIt [37] takes the deliberate strategy of providing no information about users beyond their names to "not provide group cues to activate political identity". OpinionSpace [19] takes the opposite approach of displaying users according to their answers to the values question set as described earlier. It takes advantage of the fact that liberal and conservative users often have similar answers to the values question set, resulting in closely spaced points in the 2-D space, contrary to expectations of seeing them on opposite ends of the space. This disrupts users' binary mental models and "conveys that the range of opinions do not fall along a single axis and that they are

[1]Circlejerking defined by Allison et al. [2] as "a slang term referring to the mutual appeal to and gratification of shared interests and tastes within a community"

far more diverse." With both shared and active subreddits, we build on OpinionSpace's underlying principle that revealing information about users would show that users are not as divided as they are projected to be. By showcasing non-political group memberships, users are presented with a more complicated picture about their interlocutors which we expect will disrupt the 'us' vs 'them' partisan group dynamics.

2.3.3 Exposing user information in online deliberation. A significant concern with online deliberative systems is that interactions are often between users who know nothing about each other, leading to concerns about trust and credibility of information exchanged [79]. For example, Kriplean and colleagues on evaluating ConsiderIt noted that "almost immediately after raising the issue of trust, user study participants would comment that they wanted to know more about the point author." However, as discussed above, they do not include user details to prevent priming partisan identity. In contrast, our design choice to show non-political group activity details to reduce partisan identity salience may also help to increase trust by providing individuating information. For example, Tanis et al. [74] found that, as predicted by the SIDE model, revealing individuating information about an anonymous outgroup member online increased interpersonal trustworthiness as the member is seen less as an outgroup member and more as an individual. However, revealing information about group memberships comes with multiple concerns. Firstly, it raises concerns about inadvertently revealing sensitive private attributes [83]. Secondly, revealing this information may result in an asymmetrical disclosure, where one party knows information about the other but not vice-versa. Studies, albeit on dating practices, show that even when this information is obtained from public Facebook profiles, it is typically considered deceptive and norm violating to use it [25]. Finally, this information initiates a form of 'context collapse' [44]. On Reddit, usually user activity in one subreddit is not directly visible in another subreddit allowing users to relatively freely participate in subreddits related to unpopular or stigmatized topics without it affecting their other activities (although throwaway accounts are still common) [15]. Thus, disclosing this participation information can cause real harm and harassment, especially given Reddit's known toxic participatory cultures [46]. Therefore, we carefully evaluate if and when users consent to share their activity details with others.

3 RESEARCH METHODS

3.1 Research Context

We conducted this study on Reddit users in the lead-up to the 2020 U.S. Presidential elections. Reddit is a popular social networking platform comprising of hundreds of thousands of subcommunities called subreddits. Each subreddit is centered around a topic and independently run by volunteer moderators. Although there are some commonalities, the norms and rules enforced in these subreddits may also vary significantly [9, 20]. For example, r/NeutralPolitics and r/moderatepolitics both host cross-partisan discussions but vary in how the discussions are conducted. While the former does not allow "bare expressions of opinion" and requires claims to be backed by sources, the latter has no such restrictions. Users interact with each other in these subreddits through a threaded comment system that allows users to directly reply to each other. This allows for prolonged interactions between pairs of users. Comments accumulate points (called karma) through up/down votes by other users which affect their visibility. Users accumulate karma points as well, which is the sum of their comments' karma points. A similar mechanism applies to the top-level posts in the subreddits called "submissions". Many cross-partisan interactions take place usually in relatively non-partisan subreddits such as r/PoliticalDiscussion, question-answer subreddits such as r/AskTrumpSupporters, ideological subreddits such as r/neoliberal and occasionally in partisan subreddits such as r/politics. As an indicator of the levels of partisan animosity prevalent on Reddit,

Participant	Recruitment	Age	Gender	Ethnicity	Political Orientation	Years on Reddit
P01	PM	37	Male	White	Left	10
P02	PM	19	Male	East Asian	Left	1
P03	PM	23	Male	White	Right	3
P04	PM	35	Male	White	Left	7
P05	PM	36	Male	Caucasian	Ind./Right-leaning	10
P06	Univ.	20	Male	Caucasian	Right	5
P07	PM	21	Male	White	Right	3
P08	Univ.	25	Female	Chinese-American	Left	6
P09	Univ.	24	Male	Hispanic / Latino and White	Left	7
P10	Univ.	28	Male	Middle Eastern / Southwest Asian	Ind./Right-leaning	15
P11	Post	-	Female	Black	Left	2
P12	PM	37	Male	White	Right	5.5
P13	PM	48	Female	Jewish	Left	7
P14	Univ.	23	Nonbinary / Genderqueer	(Southeastern) Asian	Left	8
P16	PM	62	Female	Caucasian and Native American	Right / NeverTrump	1.5
P17	Post	33	Male	Black	Left	1.5
P18	Post	22	Female	Caucasian	Left	5

Table 1. Demographic details of participants. We report participant responses to a short open-ended demographic survey as submitted by them. P11 did not provide age details. Recruitment channels are PM (Reddit private message), Univ. (university mailing lists) and Post (post on subreddits).

many large political subreddits such as r/The_Donald and r/ChapoTrapHouse were banned for inciting hate just a few days before our first interview. It is in this context that we studied the strategies that users engage in cross-partisan discussions and the potential effectiveness of our designs in facilitating quality discourse.

3.2 Participants and Recruitment

The participants of this study are United States residents who actively use Reddit to have cross-partisan political discussions. Participants were recruited through Reddit private messages, recruitment posts on subreddits such as r/PaidStudies and multiple university mailing lists. First, we tried recruiting by sending private messages on Reddit to users inviting them to participate in the study from a Reddit account created for this purpose; we did not get any responses. Speculating that the lack of response was due to the account being new and not trusted, we sent recruiting messages through the first author's personal account which was much older, had more karma points and detailed history. This approach was more successful, 9 out of 83 (> 10%) users to whom we reached out agreed to participate in the study. We sent recruitment messages to users who actively engaged with opposing partisans in political subreddits such as r/politics, r/AskTrumpSupporters and r/moderatepolitics. However, this approach appeared to predominantly recruit White males, likely due to privacy and safety concerns. Therefore, we turned to two other channels: university mailing lists and subreddits such as r/PaidStudies. These are both popular recruiting avenues for academic research where we could more easily identify ourselves as university researchers to

establish trust. We were able to recruit a more diverse set of participants using these approaches. The interviews were conducted from July to September of 2020. In total, we conducted interviews with 18 participants. For this paper, we exclude P15 who in her interview explained that she only lurked on political subreddits and did not participate in them. Participants were required to be (i) US residents, (ii) 18 years or older and (iii) must have participated in cross-partisan discussions to be eligible for the study. Each participant was paid with a $20 Amazon gift card as compensation for their participation in the study.

Table 1 lists the demographic details of the participants.[2] 11 of the participants identified as male, 5 as female and 1 as nonbinary/genderqueer. Participants ranged from 19 to 62 years of age, with most participants in their early twenties. They skewed mostly young, white, and male, paralleling the general demographics of Reddit users [3]. 10 of the participants were left-leaning, 5 were right-leaning and 2 were right-leaning independents. We interviewed participants in different occupations such as software programmers, university administration staff, high school teachers, census workers, undergraduate and graduate students. P12 is also a moderator of a political subreddit. Participants' experience on Reddit range from 1-15 years with a median of 6 years of involvement, and many spent months lurking before creating their account.

3.3 Data Collection

The interviews were conducted by the first and second authors. Almost all participants were interviewed using video conferencing software (except P16 with whom we conducted a telephonic interview and narrated the designs instead). The audio was recorded after obtaining informed consent and later transcribed. The median duration of the interviews was 55 minutes. Each interview consisted of two parts: a semi-structured interview (around 40 minutes) and a design probe interview (around 15 minutes). From the semi-structured interviews, we obtained rich and detailed information on their motivations, positive and negative discussion experiences, and strategies they use to participate in these discussions. In the design probe part of the interview, we shared 2-3 designs based on decategorization and cross-categorization strategies on screen and after a brief explanation of the probe, we asked for their feedback and reactions to the probe. We also specifically probed for concerns they may have had about using the interface and about others using this interface when interacting with them.

3.4 Data Analysis

Each interview was transcribed using otter.ai before manual revisions and corrections by the first and second authors. The interviews were coded using a grounded theory approach [11] consisting of both open and axial coding using NVivo software. The first and second authors independently coded the interviews (12 and 5 interviews respectively) using open coding. These codes were then combined into higher level categories using an axial coding process. The two authors met multiple times to discuss and combine these categories, and identified emerging themes around (i) motivations for participating in cross-partisan discussions, (ii) qualities of political discussions, (iii) proactive and reactive strategies adopted by participants to have good discussions, (iv) folk theories of why cross-partisan discussions are difficult to sustain, and (v) humanizing effects of the design probes and concerns around misuse. Through the course of interviews, we held weekly meetings with the research team to discuss the feedback from interviews about the designs, allowing us to incorporate minor modifications to the design probes detailed in Section 3.5.

[2]We report participant responses as is from a short open-ended demographic survey.
[3]www.journalism.org/2016/02/25/reddit-news-users-more-likely-to-be-male-young-and-digital-in-their-news-preferences/

3.5 Design probes

Currently, the Reddit interface, as shown in Figure 1a (the interface excluding the user card), does not directly provide any information about the interlocutor. Users need to hover over the username to obtain basic profile attributes such as time since joining Reddit and total karma points. To view past comments or other subreddits their interlocutor has participated in, the user has to go to the interlocutor's profile page by clicking on the profile icon. Through our design probes [21], we explore alternate versions of the Reddit interface where the user has access to additional information which is expected to decategorize or cross-categorize their interlocutor. By visually showing designs containing this extra information, as opposed to asking participants to imagine such a possibility, we provide a realistic representation of this information on which to base their opinions. The aim of the designs are two-fold: (i) To understand how participants perceived the impact of the extra information on their conversations and (ii) To explore different designs based on participant feedback to build a functional browser extension. Below, we detail each component of the user card which is intended to show up when users click the 'reply' button to reply to another user's comment (as shown in Figure 1a).

3.5.1 (A) Shared subreddits. This component shows the list of non-political subreddits that both the participant and their interlocutor have recently participated in. By explicitly highlighting shared group memberships, we alert the user to the presence of cross-cutting identities which is found to have a calming effect on partisan hostility as described in Section 2.2.1. The subreddits were be ordered such that smaller subreddits were shown first since group size is negatively associated with affinity towards the group in online communities [36].

Feasibility Analysis. Displaying shared subreddit memberships to interlocutors will be beneficial only if they actually share subreddit memberships. This concern is especially significant now as political science research suggests conservatives and liberals on average make different choices on even non-political decisions such as coffee choice and fast food consumption [16]. This may result in few common subreddit memberships between outpartisans. Therefore, to evaluate the reach and thereby effectiveness, we estimate the prevalence of shared non-political group memberships among users who engage in cross-partisan discussions on Reddit using publicly available data [5].

First, using a simple heuristic from prior work on Reddit [59], we identify users who are left or right-leaning based on their activity in left and right-leaning subreddits. First, we identify r/politics, r/Liberal, r/progressive as left-leaning and r/TheDonald, r/Conservative, r/Republican as right-leaning subreddits. Then, we classify users as left-leaning if (i) they comment in more left-leaning than right-leaning subreddits (ii) the mean karma points of their comments in left-leaning subreddits is higher than their score in right-leaning ones and (iii) their mean karma score in left leaning subreddits is greater than 1. Likewise, we identify right-leaning users. [4] Then, using these user classifications, we identify all distinct copartisan and cross-partisan interlocutor pairs in 277 political subreddits (previously identified by [60]). For each pair, we identify if they both participated in a common subreddit within the last 3 months, while excluding the 277 political subreddits and the default subreddits[5] from consideration. We find that, in an average subreddit, 44.26% and 51.94% of all cross-partisan and copartisan discussion pairs share at least one common non-political subreddit. These percentages are encouraging because (i) in an political average subreddit, about half of all discussion pairs share a non-political subreddit indicating that showing shared subreddits is a viable option for a sizable population of interactions, (ii) the difference between copartisan

[4]We classify 1,223,229 users as left leaning and 367,363 users as right leaning. We cannot identify the political leanings of other users using this approach.

[5]Until June 2017, Reddit users were automatically subscribed to these subreddit which are amongst the largest on the site.

and cross-partisan percentages, although statistically significant, is small enough to suggest this difference may not significantly exacerbate outparty differences, although experimental work is required to estimate the effect of *not* observing shared subreddits. However, these results also highlight an important limitation of this design: less than half of all cross-partisan interaction pairs can be shown the shared subreddits component.

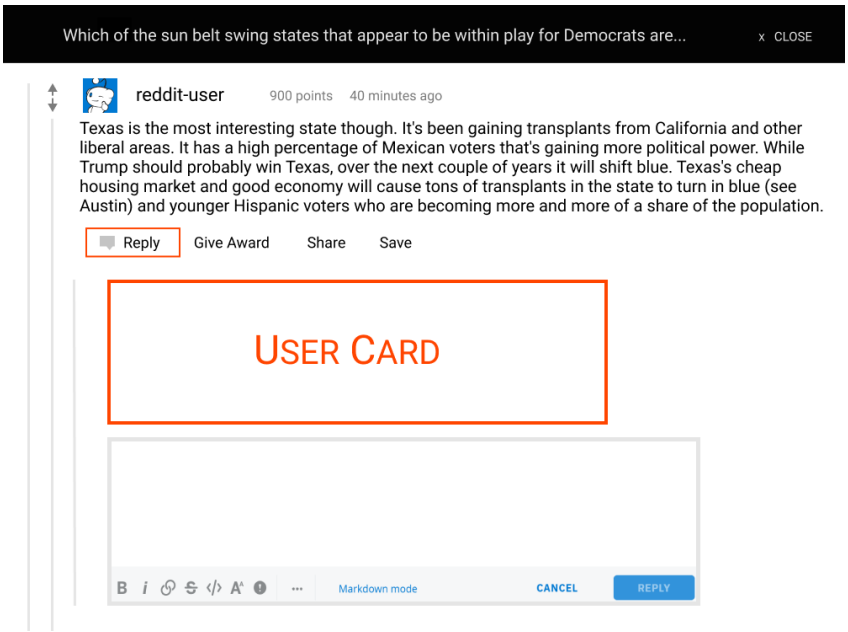
3.5.2 (B) Active subreddits. This component shows the list of non-political subreddits that the interlocutor has recently participated in, excluding the "shared subreddits". By explicitly highlighting the interlocutor's varied interests and identities based on their activity, we aim to reduce hostility through decategorization as described in Section 2.2.2. Again, these subreddits were be ordered such that smaller subreddits were shown first.

Feasibility Analysis. Displaying active (and not shared) subreddit memberships will be beneficial only if they actually participate in other subreddits. Redditors are known to create multiple throwaway accounts to provide added anonymity, especially when discussing contentious issues [39]. If users predominantly use throwaways when talking about politics, then there may be few other subreddits to display to interlocutors. Therefore, using a similar approach as earlier, we calculate among users who participate in political subreddits in a given month, the average number of nonpolitical subreddits they participated in the prior three months. We find that the left and right leaning users active in political subreddits in 2019, on average, engage in about 23 and 20 subreddits respectively in the prior three months, providing evidence that this design is indeed feasible given current user behavior data.

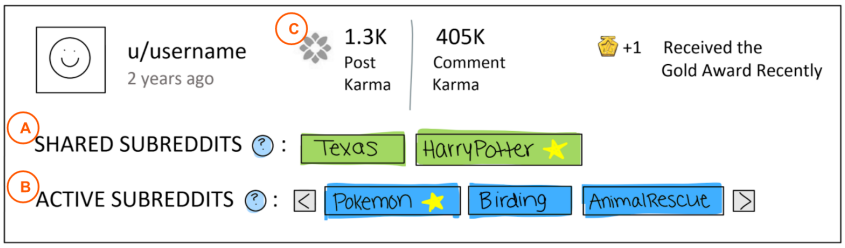
3.5.3 (C) Karma points and awards. This component shows the karma points and awards earned by the interlocutor. Though unrelated to decatagorization or cross-categorization, karma points may have potential to improve conversations by providing an indicator of trust or reputation bestowed on the user by the Reddit community. This feature was designed based on feedback from ConsiderIt where their study participants expressed difficulty in evaluating the trustworthiness of claims put forth by other users about whom they knew nothing [37]. Highlighting awards and karma points could present one way to highlight trust without giving away partisan cues about the user.

3.5.4 (D) Comment highlights. This component highlights top comments posted by users in non-political subreddits based on karma points. By providing examples of top non-political comments by the interlocutor, we aim to showcase their positive behavior in other subreddits indicating that they have multiple interests apart from their politics. Along with the active and shared subreddits, comment highlights provide deeper insights into not only where they participate but also how they do so in the other subreddits. A more discrete version of comment highlights is the star-shaped link in the active and shared subreddit boxes which links to a top comment (above 50 karma) posted by the user in that subreddit.

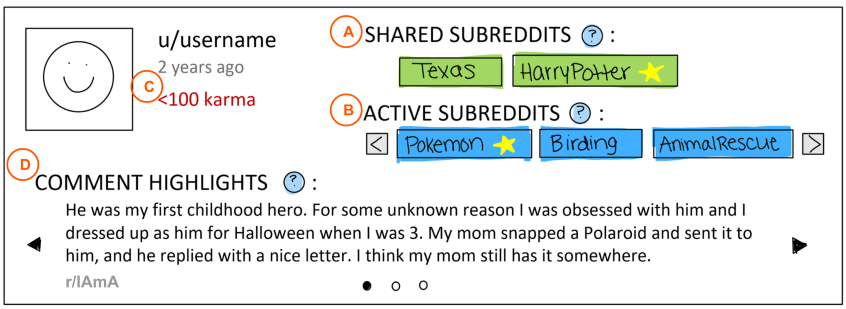
Design evolution. First, we conducted interviews using one design (Design A, Figure 1b). During the first few interviews (P1-P3), participants suggested providing cues about not just *where* users were active but also *how* they behaved in those places. This feedback resulted in us evaluating additional designs that made visible more details such as "comment highlights" in Design B (Figure 1c). We showed Design A to all participants and Design B to P4-P18. We also developed other largely similar versions of these designs aimed at reducing the size of the user card by moving the placement of karma points, showing shared and active subreddits on the same row and linking to a highlighted comment rather than displaying full text. All designs featured shared and active subreddits, the primary focus of our research.



(a) Example of a comment on the Reddit interface with our user card. The user card would appear when users click the reply button to type a reply.



(b) Design A: The user card shows active and shared subreddits as well as karma points and awards.



(c) Design B: In addition to components in design A, the user card shows comment highlights.

Fig. 1. User card designs

As a cautionary tale, we note that in our case, the initial mode of recruitment (private messages) resulted in mostly White/Caucasian male participants (P1-P7) who did not have major concerns about their information being made more visible in the user cards. However, as we interviewed a more diverse set of participants recruited through other channels later in the study, concerns about revealing information became clearer. We caution that designs such as ours that highlight user information needs to be carefully evaluated for their effects, especially on members of disadvantaged groups early in the design process.

Target demographics. A significant advantage of these designs is that they are not explicitly political; users would simply see the non-political activity of other users. Therefore, an extension built using these designs can be marketed as a fun tool that helps users learn more about others, which we expect will help diversify the kinds of users who install the extension. By positioning it as a general-purpose fun tool, we anticipate that all users, not just the ones most motivated to improve their discussions, will use the extension. However, because of the nature of these designs, we expect it to be less effective on extreme partisans whose non-political subreddit membership is likely stereotypical. Further, the extension is not expected to reduce hostility expressed by individuals who are determined to be hostile, rather it is a subtle intervention aimed at users who engage in cross-partisan interactions in earnest.

4 FINDINGS

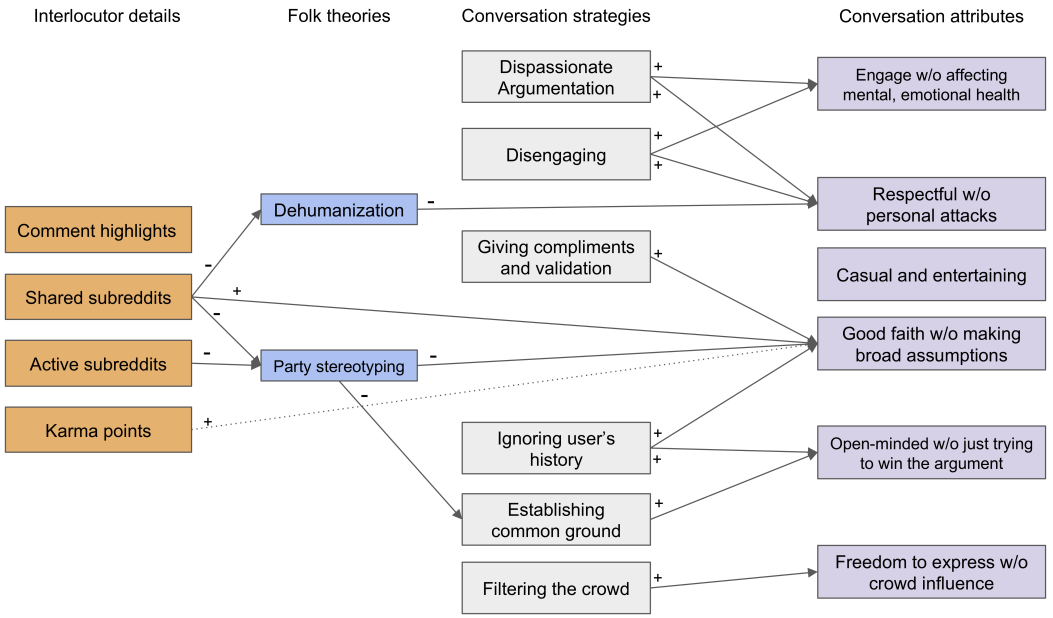


Fig. 2. Summary of findings showing the relationships between good conversational attributes and user strategies employed during the conversation, folk theories and interlocutor details made available through design. The arrows indicate the directionality of the relationship between the entities, and the signs (+/-) indicate whether the relationships were positive or negative. For example, establishing common ground increases the odds of having open-minded discussions while an increase in dehumanization decreases the odds of having respectful interactions. The line from karma points to good faith is a dotted line since karma is a weak/basic indicator of good faith.

We organize our findings as follows: First, we detail the qualities that participants seek in a good cross-partisan political discussion (rightmost in Figure 2). Next, we highlight strategies that participants adopt to improve the chances of experiencing these good qualities in their discussions (center-right, in grey). Then, we detail two folk theories–dehumanization and stereotyping–that participants attribute to the many bad conversations they have despite following these strategies (center-left, in blue). Finally, we explore how the user information embedded in our designs may help overcome dehumanization and stereotyping but may also lead to other concerns (leftmost). The (+) and (-) signs in Figure 2 indicate positive and negative relationships between the entities. For example, establishing common ground increases the odds of having open-minded discussions while an increase in dehumanization decreases the odds of having respectful interactions.

4.1 What is a 'good' cross-partisan political discussion?

When asked what they considered to be a good cross-partisan interaction, participants described two kinds of interactions: (i) serious deliberative discussions on political or policy issues and (ii) casual conversations for entertainment and banter. Interestingly, many participants reported engaging in both kinds of conversations depending on their mood or time constraints. We describe these conversations in detail below.

4.1.1 Serious deliberative discussions. Most participants expressed that they were looking for some form of serious deliberative discussion. Many of the specific attributes they looked for in such conversations directly mapped to the deliberative ideals of mutual respect, reasoned arguments and the freedom to express without coercion (such as crowd influence).

Respectful without name-calling and personal attacks. Most participants expressed that they aim for conversations to be polite and respectful without devolving into personal attacks.

The bad conversations start off right away antagonistically, you'll have like a Trump supporter or a liberal supporter like basically just start off by saying nasty ad hominem attacks about the other side, these are already like non-starters like you're not going to get anywhere. - P01

Listening with an open mind without simply trying to win the argument. Most participants entered cross-partisan discussions without expectations of changing others' views. Instead, they looked for conversations where their interlocutors were simply open to acknowledging some of the issues that they had raised.

For instance, [pretend] you're pro-Trump. But I made a point that you can't find anything to disagree with about. Can you actually say that? You know, while I support Trump, you actually have a valid point on this particular issue. So, being willing to listen is to at least consider what the other person is saying, which does require listening, is huge. - P16

However, participants indicated that many conversations are not open-minded exchanges of ideas but rather interlocutors simply trying to one-up each other. Thus, some participants do not even look for open-minded users, instead, they use the conversation to explore an issue. For example, P12 recounted instances where he argued with others "for the sake of just understanding that idea". Few of our participants actively looked to change others' views. Still, they reported that instances where they changed others' viewpoints were uncommon.

Good faith without making assumptions. Participants look forward to having good-faith conversations with others–conversations where everyone has good intentions, engages in earnest, and refrains from making assumptions of others.

[A good conversation needs] understanding that each person participating has experiences that you might not be able to relate to or like, language is really imperfect... understanding that like everyone is

like trying to do right by their communities and families. So even if like we can't understand what those obligations look like, they are "good people"... - P14

However, participants noted that in many of these conversations, people quickly make assumptions and judge others without giving them a chance to explain their beliefs.

Informative without unverifiable claims and party talking points. A prime motivation for almost all participants to engage in cross-partisan political discussions is to learn about opposing viewpoints and contribute alternate perspectives. Similar to Semaan et al.'s [66] findings, most participants in our study explicitly acknowledged the effect of filter bubbles or echo chambers on their own beliefs. They stated that they actively try to engage in cross-partisan political discussions to gain alternate perspectives.

I enjoy talking with [conservatives] because they'll see an article and see it completely different than the way I see it. It's a curiosity for me... And it's good for me to know that they exist, and not just this little bubble that I'm in. - P04

However, in many conversations, participants noticed that the interlocutors simply regurgitated party lines or spread debunked misinformation without researching on their own to understand the issue.

When the person is not willing to debate facts, when they start spewing basically talking points, talking points that are disputed, talking points that aren't related, talking points that don't make sense... then that's a pretty good indication it's not going anywhere. - P17

Freedom to express without crowd influence. Almost all of our right-leaning participants and many left-leaning participants described how their comments are often heavily downvoted or heavily replied to by many users (dogpiled), which overwhelms them.

Reddit, a lot of it is primarily liberal. So it's like, if you want to come with any conservative opinion whatsoever, you're probably just going to get mobbed on and, you know, for every 100 people that mob on you, there might be five actual discussion points in there. - P05

However, participants sometimes do take into account the feedback they receive from others, especially copartisan feedback. As P13 described, receiving downvotes or multiple replies does prompt her to reflect and question her positions on the issue.

People will reply vehemently... I'm really surprised when it's more than two people... It makes me wonder whether or not my position on that topic should be that position. Do I change my mind? No, not necessarily, but the thought is there and that's important too, because you have to constantly question your own thoughts. - P13

Engaging without affecting mental and emotional health. Many specifically described the toll that some conversations had on mental health. Participants report that cross-partisan conversations often involve a lot of work and mental effort, the after-effects of which may continue to linger through the day. Even conversations that do not veer into name-calling or character attacks sometimes leave participants frustrated and exhausted.

The fear is that like, it'll consume the whole day. I'll be thinking about something politically... and I'll just keep talking about, thinking about like, something political and get worked up about it. - P07

However, some participants are able to ignore views that they dislike, or quickly move on to lighter content to decompress, as P11 put it, "continue to scroll to find a cute puppy". Others, like P2, are resigned to the fact that being exposed to objectionable views is the cost of having a cross-partisan conversation.

But it's those little prejudices that people have that bother me, but while I do feel bad reading them, I don't think I am necessarily upset because of it. Because the main reason why I'm on there is for a political discussion. - P02

4.1.2 Casual and entertaining conversations. Participants explained that they were on Reddit primarily to have fun and entertain themselves. They did not use Reddit for political discussion alone and most actively engaged in other relatively non-political communities such as r/DIY and r/Makeup. They saw their participation in political conversations as one among many other leisure activities they engaged in on Reddit. In fact, some participants explained that many of their political conversations were incidental and stemmed from casually browsing through their home feed. They were not actually trying to engage in the conversation deeply and would typically quickly comment and leave.

I kind of will just comment whatever, not really trying to seek out [conversation] because also, once you do get a productive conversation going, it takes a lot of energy, it can take a lot of time... I don't know if I have the stamina for it all the time. - P17

Sometimes, participants engage in more casual political subreddits such as r/PoliticalHumor and r/PoliticalCompassMemes. However, many find discussions in mainstream discussion subreddits entertaining as well. P09 described how he uses Reddit most heavily when he is bored at work, and primarily looks for entertainment when participating in political discussions. Below, he described one such discussion where a conversation in r/politics devolved into a conspiracy theory.

One of the funniest ones I ever remember reading was a pizzagate-like thread of comments. It was just hilarious. Because that was a conspiracy theory and then lots of people branch off and like, I don't know, it was so entertaining... Like watching some people just put two and two together on things I could have, like, never thought twice. That's some high entertainment value. - P09

Participants also mentioned that even within more serious discussions, someone may post a witty rejoinder or a funny meme which makes the tone of the conversation fun and casual. They also pointed to how some political subreddits have dedicated discord chat servers and occasional free talk threads, allowing users to have casual discussions unrelated to politics. In certain instances, participants indicated that normatively anti-social behavior such as trolling and making others angry were also fun activities to take part in on political subreddits.

If I am in a really bad mood, and I'm just out to, you know, troll up a storm, then what I think of as good [conversation] is when I get someone's goat, when I make them very viscerally angry, and they keep responding, and I can tell they don't want to respond, but they have to respond and that's when I've got them! - P10

This poses a direct contradiction with one of the most commonly cited motivations—to have respectful deliberation. It is important to note that the same participants who seek entertainment in political discussions also participate in more deliberative political discussions. The kind of conversations in which participants choose to take part depends on a wide range of factors such as mood, time constraints and current events.

Frequency of "good" discussions. Most participants reported that they participated in at least a few political discussions that they felt were good and satisfying. However, these occurrences were rare. Many long-time participants reflected that their conversations have turned angrier and devolved more into name-calling over the past two years. However, many participants characterize engaging with the other side as a form of civic duty, something that is difficult but needs to be done to deeply understand issues affecting the country. Therefore, to navigate these conversations and increase the odds of having a good interaction, participants have developed multiple strategies to select where, who and how to talk to cross-partisans which we detail in the following sections.

4.2 Strategies adopted prior to engaging in cross-partisan discussions

4.2.1 Choosing where to have the conversation. Many participants reported taking part in multiple political subreddits and carefully curating the subreddits that they subscribed to, considering the

quality of discussion, member composition and level of moderation in conversations in those communities. Some participants completely avoided large generic subreddits such as r/news and r/politics and instead participated in relatively smaller niche political subreddits such as r/tuesday which is a relatively small center-right subreddit whose participants are derided as RINOs (Republicans In Name Only) for not being Republican enough and r/moderatepolitics, a moderate-sized non-partisan discussion subreddit where they could have more nuanced conversations. [6] Note that these niche subreddits are not homogeneous partisan groups. They were relatively much smaller, well moderated, and frequented by members who similarly value cross-partisan interactions. P05 explained why he customizes the subreddits he participates in:

The reason why I follow some of these particular subreddits is just because people seem a little bit more reasonable in how they respond. You know, we probably both know that Reddit is primarily liberal, just in general. And you kind of have to go to specific subreddits if you want to get say like, right-leaning information or commentary. But some of the subreddits like r/Conservative and, you know, now banned, r/TheDonald, they're just so far over there. And the quality of discussion, in my opinion, is very, very low. - P05

Through experience, some participants claimed to understand how their comments will be received depending on the type of subreddit in which they participate. A few others explained that they participated in subreddits that only partially aligned with their views which allowed them to have disagreements knowing that there was also some common ground.

The r/Neoliberal one is a good one for me, just because I do dissent somewhat from some of the things they believe, but I also have a lot of common ground. So it's a space where I can have a lot of discussions with people who write, you know, at least they have similar moral frameworks, similar sort of ideological frameworks, even if some of the actual practices diverge a little bit there. - P06

Even within subreddits, a few users are selective about which threads to engage in. For example, P04 explained that he recently started engaging in the open talk discussion threads on r/tuesday, rather than topic-specific ones. He reasoned that most of the subreddit "regulars" hung out there and were able to have deeper conversations since these threads do not usually get upvoted enough to show up on people's homefeeds and attract widespread attention from casual users. Generally, participants attest that identifying the right space to participate in tremendously affects all aspects of the discussion.

4.2.2 Choosing who to talk to. All participants said that they viewed only the text of a comment by a user to decide whether to engage with them. Many participants described having an intuitive sense of how the conversation was going to unfold based on their reading of a users' initial comments. Some said that they could understand the 'personality' of the user by reading between the lines to make quick judgments about whether to talk to them.

I can usually tell from the first comment–and I assume other people could as well–about the tone of the discussion with this person... Most of the time, it's just like, I know, that's going to be a bad convo, and that this is going to be more reasonable... I think I think it's probably about 90 or 95% of the time, the first comment generally identifies how the conversation is going to go. - P05

Many participants also explained that they try not to enter into discussions with users who use profanity or strong emphasis words such as 'obviously' or 'clearly' when making suppositions on a topic. However, some participants also recalled times when they deliberately chose to engage with users making such statements when they were in a combative mood.

[6] r/tuesday and r/moderatepolitics had about 12,000 and 50,000 subscribers respectively at the time of conducting the study.

4.3 Strategies adopted during cross-partisan discussions

4.3.1 Establishing common ground and posing questions. Participants recognize the current contentious political climate and are extra careful in how they communicate in cross-partisan discussions. Most participants reported that they typically start by signaling common ground with the user, highlighting parts of the argument that they agree with politely and respectfully. Then, they detail aspects that they disagree with while remaining extremely deliberate about how they frame their critiques, often posing them as questions. Many noted that they sometimes rewrite their comments multiple times to ensure that their views are conveyed accurately but without offending the people to whom they are talking. For example, P13, a high school teacher, described how she communicates with those she disagrees with.

I find that when I'm super careful about how I engage somebody whose opinions I differ with, the more careful I am, the better the conversation goes... rule number one when you have a parent-teacher conference is "this is somebody's kid, say something nice." So for online, it's what do you agree with? What did this person say that you wholeheartedly agree with? and start from there. And then after that though, don't attack, [instead] question... - P13

Thus, by establishing common ground and approaching conversation partners without a heavy gavel, users aim to signal that they are **open-minded and reflective**.

4.3.2 Giving compliments and validating the interlocutor. Some participants, upon sensing that a conversation has turned for the worse, typically give one last shot at reviving the conversation by explicitly complimenting or validating the interlocutor, as a sign of good faith. P18 explained how they sometimes try to correct the course of the conversation.

I think if someone's aggressive, but you can kind of sense that they do have the ability to have a better conversation, I think just being nice, I think not playing into their tricks, even validating them in a certain way helps... I love to say well, I agree with that. But I also have these things that I believe in and this and that, so I think actually validating them a little bit... you kind of show them your cooperation, they might actually come out to be more cooperative and I've seen that happen. - P18

Thus, by acknowledging and validating the interlocutor, participants signal **good faith** to the interlocutors.

4.3.3 Dispassionate argumentation. Many participants explicitly aimed to keep a calm and dispassionate demeanor when interacting with cross-partisans. Knowing that disagreements tend to engender strong emotions, participants keep the conversation on topic by focusing on the facts, providing arguments, and trying not to react emotionally to the interlocutor's arguments. For example, P04 described a particularly difficult conversation with a right-leaning user who argued that reports of police brutality in the US were overblown:

I always try to start off with a very dispassionate response. And try to back up my claims with as much fact as possible and try to keep feelings out of it as much as possible, leave my worldview out of it as much as possible because we clearly don't share the same worldview. So I'm never going to be able to win that person over with that aspect, but just try to make it dispassionate. - P04

However, he conceded that being dispassionate on topics like the Black Lives Matter protests is especially difficult and instead, chose to disengage altogether. In other instances, once participants sense that a user is becoming emotional in their replies, they swiftly disengage or concede the argument before it (potentially) devolves. For example, P12 said:

There'll be times when I just stopped a conversation because someone's emotional. And I'll just concede the argument. There's no point in pushing somebody into a character attack when they're just getting emotionally invested in the argument. - P12

Thus, users aim to remain dispassionate in their conversations to **maintain their own mental health and to prevent the conversation from potentially devolving into name-calling and character attacks**.

4.3.4 Avoiding looking at the user's profile unless the conversation goes stale. Many users actively avoid learning more about their conversation partners by refraining from viewing their profile details such as karma points or past comments unless the conversation goes awry. By ignoring other possibly disturbing details about their conversation partner, participants focus their attention squarely on the argument that the person presents, not biased by their past opinions on other topics.

Normally, if I know that the other person that I talked to is a huge racist, or a sexist, or has some very, you know, skewed perspective on the world, then I would immediately want to stop talking to them... so I tend not to read the other person profile. I just try to, you know, discuss the topic with them, just that topic and nothing else... I don't want to know about that person, other than the things that are relevant for that discussion. - P02

Others who do view past comments express doubts about whether knowing more about the user helps or hurts the conversation. For example, P04 was concerned about whether knowing a user's position outside the topic might prejudice him in the conversation.

I have [viewed past comments] in the past, especially with a name that I don't recognize, just to get an idea of what I'm getting myself into. But at the same time, I almost feel like that kind of almost prejudices me. And I almost want to have a narrowly scoped discussion that doesn't have the baggage of previous discussions or previous outside-of-this-subreddit's discussions. - P04

Others, like P14 felt like knowing more about a user makes having a conversation with them difficult since the distance between their worldviews becomes more apparent. Looking at user profiles, most participants viewed karma points not as a predictive indicator of whether the user would be a good person to talk to but instead more as an explanatory variable when a conversation goes awry to make sense of the user's behavior.

Thus, by not viewing the user comment history or karma points, participants essentially try not to learn more about the user, ensuring that they discuss in **good faith and with an open mind without making assumptions**.

4.3.5 Filtering the crowd. Many participants recalled instances where their reply notifications "blow up" when multiple users angrily replied to their comments. In those instances, participants typically put on "social blinders" and focus on replying to only a specific individual.

Another thing I might do if I had a bad argument, but I liked one of the other people in the audience, and then everyone else is a bit [much]... I might sort of put like, social blinders on, and just tag that one person over and over again, to make it clear I'm only talking to them. Or I might continue the conversation in the chat box with them, Reddit has chat now, and continue in the DMs. - P10

Unrestrained by the topic or other users in the subreddit, P16 found that DMs allow for users to switch topics and be more open. She noted that "it just seems to me though, that direct message allows for a level of intimacy and being real." In other instances however, multiple participants reported that they have been directly targeted or harassed by others through DMs.

4.3.6 Disengaging from the conversation. By far, the most common reaction to a conversation that regresses into a personal attack or becomes combative is to disengage and exit the conversation.

I don't have to sit there and have somebody be ugly to me. That's not what I'm on the Internet for. I'm on the Internet to have fun and to be educated and not to be harassed. - P11

Some participants use more stringent methods to disassociate themselves from the conversation by deleting their comments, reporting to the moderators, blocking the user and in rare cases,

unsubscribing from the subreddit. It is important to note that disengaging is not a last resort action that participants take, oftentimes, disengaging is the first action that they take. Thus, to **safeguard their own mental health and to shield themselves from personal attacks**, participants simply disengage and walk away from the conversation.

4.3.7 Counter strategies. Not all strategies employed by the participants are conciliatory or aim to further the discussions. In some instances, participants said that they would counter by using aggressive or condescending language.

if I'm feeling petty, it's not like the right thing to do, but I like call them out in kind of a condescending way, I don't like using insults and things like that. But if I do want to be petty, it'll be more like, yeah, condescending or rhetorical questions, lighter, but still, I know, I shouldn't be talking like that. - P08

In other situations, recognizing that the user they are talking to is angry, some users try to make them angrier.

Normally, when they're really mad and they go out of their way to like, target me. I normally just like, take the piss, you know, I kind of try to make them more mad... I don't confirm their prejudice. I just go, you know, oh man, look at this guy...haha... It's kind of stuff like that. - P07

Others described using some of the tactics described by Jhaver and colleagues [33] such as identity deception and sockpuppeting to counter hostility.

Do these strategies work? *Sometimes.* Most participants acknowledged that while they do employ many of these strategies, the most effective approach in dealing with volatile conversations is to leave. Many recalled instances where they've tried to course correct a conversation only to make it worse. For example, P13 said:

I tried to engage once with somebody that vehement, and they just were, they just attacked. It was like, you know, it was like getting a text that's like, all caps from your mom. And it's just, you know, who needs that? So I'll just drop it. I don't reply. I just let it go. - P13

Most participants explained that it is best to find another conversation to participate if their current conversation became worse. As P6 put it, "when you invest a lot of energy into what is effectively an online discussion, it can sometimes feel like shoveling money into a fire."

4.4 Party Stereotyping and Dehumanization: Folk theories on what affects their conversation

While we did not specifically ask participants why they thought their strategies did not always bear fruit, many participants had unprompted explanations of their own, specifically attributing party stereotyping and dehumanization as a cause for concern in cross-partisan discussions. It was revealing (and frankly surprising) how well these folk theories matched with our cross-categorization and decategorization attempts through design.

4.4.1 Party Stereotyping. Some participants attributed certain conversations going awry to stereotyping along party lines. In their experience, some users were quick to judge them as extreme liberal or conservative and project on them, what they perceive to be the typical characteristics of the group. P07 explained one such instance:

I think the worst one is where like, they kind of view you as the representation of like the right-wing or something. I'm not very conservative, but it's annoying when people are like, oh, you religious conservatives. Like, I'm not very religious. I'm not very conservative, they assume that like, you represent the whole like, you know, straw man of the entire wing - P07

In other cases, participants were concerned about how a copartisan user supporting a position held by the participant may speak up for them. However, in doing so, they may provide reasons that are incongruent with the participant's own reasoning.

You end up with the problem of sometimes someone will say something as if he's speaking for you. But really, it's like, No, no, don't put me in there. [copartisan would say] "And that's the issue..Republicans, Black people. And I'm sure everyone else here [agrees with me]", please no no noooo! We are not the same, though. - P06

Therefore, party stereotyping erases differences between individual group members (both ingroup and outgroup), leading users to **make broad assumptions of each other and affecting the ability to build common ground**.

4.4.2 Dehumanization. Contrasting with face-to-face interactions or interactions with people they personally know on other social media, many participants **attribute personal attacks** to dehumanizing effects afforded by anonymity on online platforms like Reddit.

Especially the anonymity that Reddit has, it's very easy for you to forget that that's a real person on the other side or for other people to forget that's a real person on the other side, you just start like throwing vitriol and people are just like, non-caring, like, will use any type of language to try and get their point across. And it's like, hey, I'm a human being, let's be at the very least cordial, we don't have to agree, but we should probably not try to like kill each other with words. - P11

Either through personal experiences or subreddit rules or by reading the 'Reddiquette' [7] which urges users to 'remember the human', many participants recognize the need to view other users as human beings instead of a username on a screen.

I'm somebody who grew up with the internet evolving. I didn't start it when I was a kid. You know, I didn't have a phone in my hands until I was in my 20s. So I still go into every online conversation the way I would a real conversation. I'm constantly remembering that there's somebody on the other end. I'm consciously like this. I really pay attention to the words on the screen. - P13

While participants understood the importance of remembering the human, they found it difficult to practice it online without other visual or auditory cues. Most participants felt that knowing more about the user and their interests would help view them as more complete human beings rather than just as someone who has strong political opinions. For example, P06 explained that the users he talks to online are strangers and that knowing more about them would humanize them:

It would be cool to know what kinds of stuff the other person's into, and just to maybe not put a face to it, but maybe, you know, at least see some additional humanity behind what is otherwise a username and text. - P06

However, many of the participants who acknowledged the importance of 'seeing the human' remained deeply skeptical of knowing too much about their conversation partner for fear that extra information may distract or bias the conversation. For example, later, when asked if he would like to know more about users he talks to, P06 said:

In a sense, I don't want to know very much about the person other than that they are a good partner or conversationalist or whatever... I wouldn't want to know anything about the person, their race, I wouldn't want to know their gender, I wouldn't want to know shit... I think beyond including resources that clue people into someone being a good debate partner, the other information becomes more so distracting or brings about expectations that will guide the conversation in a way that is not based on the substance of the argument itself. - P06

Thus, many participants appear to navigate the following paradox: knowing too little, you risk dehumanizing them. Knowing too much, you risk the integrity of the conversation—and usually, participants lean toward minimizing the additional information they know about the user.

[7] informal norms that users are urged to subscribe to, https://www.reddithelp.com/hc/en-us/articles/205926439, "Remember the human. When you communicate online, all you see is a computer screen. When talking to someone you might want to ask yourself "Would I say it to the person's face?" or "Would I get jumped if I said this to a buddy?"'

4.5 How do users consider the extra information provided by the designs?

4.5.1 Shared subreddits.

Potential for humanizing users. Many participants stated, often enthusiastically, that viewing shared subreddits on the user card would remind them that there was a real person, a human being, on the other side of a conversation. For P08, shared subreddits would make them feel more connected to the user who is otherwise just a random stranger, and would likely reduce anger and negative emotions.

I think this would be very humanizing. I think you can see what kind of, you know, interests they have on Reddit outside of politics and the conversation that you're having... if it's happening in a negative or a politically charged conversation... you know when you're talking with someone anonymous, you can be a lot ruder, a lot more condescending, and there's not really consequences to it, but when you see this, I think for me, it would reduce my anger or my negative emotions. - P08

Potential for reducing stereotyping. Other participants explained that highlighting commonalities could help bridge the gap and see the person in a (relatively) more positive light.

I think [shared subreddits] is helpful because outside of the political spectrum people do have common interests. So my feeling might be, well, okay, maybe this person is not so bad. They like technology, they share the same interests in sports. - P17

Potential for fostering good faith and common ground. Many participants felt that viewing the subreddits they shared with another user would help establish for themselves some common ground with the user. They explained that in conversations that get particularly heated, knowing that they share a common interest would help to build some goodwill.

I think that could sponsor a little more goodwill among people, like, even if you have two people that are vehemently arguing with each other and calling each other–you know, flipping each other off verbally–if they find out that, oh, you have an ATV too, or a four-wheeler and you'd like to go out, it could sponsor a little bit more goodwill, which I think could ultimately lead to better conversations, for sure. And it's a good idea. - P05

Concerns. Some participants were concerned that there might be few instances where the users share common subreddits, however, our data analysis (in Section 3.5.1) revealed that a sizable number of cross-partisan and copartisan pairs participate in at least one common non-political subreddit. Also, a few participants explained that they would be inclined to look at the user card only if it was someone that they recognize or have spoken to earlier. They thought that this information would be less useful for one-time interactions.

4.5.2 Active subreddits.

Potential for reducing stereotyping. As expected, some participants explained that knowing the other subreddits in which their interlocutor participates would help reassure them that the person is not fixated on politics and has other interests as well.

[Showing active subreddits would help because] that'll tell me if you're not stuck in a particular way, that they do have other interests that could influence their thought process. - P17

A few participants liked that they could quickly get an idea of the kinds of subreddits in which the user participates. They explained that currently, they needed to scroll through a reverse-chronological list of their past comments to get a sense of where they participated.

Concerns. Participants expressed two significant concerns with the active subreddits component. First, some participants felt that some active subreddits could present them in a negative light. They

feared that when others view that they participated in a fun subreddit such as a meme subreddit, they may not take their political arguments seriously or worse still, use their participation in those subreddits to discredit their arguments.

[Would not like active subreddits because] I don't want it to be like, Oh, this person's trying to describe to me economics and they browse like, I don't know, but just the Jojo subreddit all day. You know, it's a very easy path for like judgment, I guess. - P07

The other concern was that, in its current form, the active subreddits component simply displayed too much information about their activity on the site. Many participants suggested that providing a way to customize the subreddits shown on their card or to opt-out of displaying the active subreddits will allay these concerns.

When you're first reading a comment from people it's like an interaction at a bar–you want to give them enough so that they come over, but you don't want to sell them the house. - P13

However, another factor that might compound these concerns is that this design when deployed as a browser extension would result in an information asymmetry, users may not even know that their interlocutor is using the extension to view information about them. This is a serious concern that we expand on in Section 5.4.4.

4.5.3 Comment highlights. Some participants liked the idea of viewing different dimensions of a person based on their top comments in other subreddits. Some even recounted past comments that became viral or were gilded (awarded Reddit gold) by other users which they would be proud to highlight on the user card. However, many participants raised two major concerns. First, participants were concerned that comment highlights based on karma points could produce a biased view of the person and cause easy judgment. They explained that most top-voted comments were either extremely opinionated, controversial, or partisan which might provide fodder for more conflict.

If someone is passionate about one thing and not passionate about the other, or somebody could have an extreme opinion about one thing and not about another. So you would see the most extreme thing, maybe, as their top comment, and then now you get to judge people on their most extreme opinion. I don't think that would be a very good idea. - P07

Others noted the comments would likely distract them from the actual conversation or may contain outdated beliefs which may color the viewer's opinion about them. Another major concern was that participants expressed feeling self-conscious about the information revealed in the comment highlights. It is telling that all four participants stating this concern were either female or genderqueer (P8, P13, P14, P18). They felt that the comment highlights made their comments more public and their profile more open to scrutiny. P14 expressed that they would likely choose how they word their comments carefully because of the increased visibility. P18 explained that she liked the way past comments were structured on the Reddit profile page, a simple list of past comments in reverse chronological order. It afforded her privacy by not being very organized or accessible.

[I like the profile page] because I know that it's not necessarily that open and always accessible and when people are touching on touchy topics and they are expressing themselves, [they] might want to keep some sort of an anonymity, I feel like having things more presented in a way that shares more information could actually be a problem. - P18

Further, P18 expressed concern that her views on one topic may be used against her when she is discussing other issues and said that she would likely have to make throwaway accounts to prevent users from connecting issues. Similarly, while P13 did not express specific concerns about the design, she had earlier described a prior experience where users racially abused her after finding out that she was a Black woman from a photo she had previously uploaded. This component likely exacerbates these concerns by increasing the visibility of specific comments.

4.5.4 Karma points and awards.

Potential as a basic/weak good faith indicator. Many interview participants expressed that it did not matter if their interlocutor amassed high karma points and awards, as they used it not so much to determine if the person posted quality comments, but to simply indicate if an interlocutor was a troll.

Concerns. Many participants explained that high karma points only indicated that the person makes good jokes or puns and it said nothing about the quality of someone's views. This somewhat lukewarm response to karma points is in line with Massarani's observation that while Redditors place some value on karma scores, they are also suspicious of users with very high scores [47]. Almost all right-leaning participants pointed out that since Reddit has more liberal users, the karma points and awards usually only reveal how liberal the user is and therefore might bias the conversation. Hearing this initial feedback, we converted our karma indicator to display only if they had less than 100 karma for use as a very basic good faith indicator. Later participants told us that this information, coupled with information on the age of the account, was enough to identify troll behavior.

5 DISCUSSION

5.1 Education vs entertainment in cross-partisan discussions

We find that participants engage in political discussions both to educate and to entertain. Depending on external factors such as time constraints and outside news cycles, the same user may engage in relatively serious political discussions or may casually peruse the site and join in casual banter, satire, and trolling. This finding is also in line with past work that shows that trolling behavior is context-dependent and not an immutable individual characteristic [12].

The two goals may be at odds with each other. The same comments may be appreciated differently by people seeking education vs. entertainment. Interventions that aim to coach users to talk to the other side (such as [81]) might help produce comments that are more effective for education than entertainment. Participants may also be more receptive to such coaching when they are primarily motivated by education rather than entertainment. On the other hand, the two goals can also be complementary. None of our participants joined Reddit for its political content and most had significant interests in other non-political subreddit topics. By hosting something for everybody, Reddit likely allows casual political observers who happen to peruse the site for other reasons to engage in political talk. Also, many participants commented that they often switched to lighter content when conversations go awry or when they simply needed a break from a heavy discussion. We speculate that this easy access to fun and entertaining content has therapeutic effects and serves to lighten the after-effects of serious political discussions. This recuperative function is particularly important given the participants' concern about the emotional toll of these conversations.

5.2 Unintended consequences of cross-partisan discourse?

The outcomes of cross-partisan interactions, both online and offline, have been typically evaluated in terms of highly valued outcomes such as political participation and political efficacy. However, little is known about the effects of cross-partisan interactions on users' emotional and mental well-being. From our interviews, we observe that most participants were wary about the discussions' repercussions on their mental health and employed multiple strategies to negate these effects. Thus, we call on researchers to attend to the psychological effects of participating in these discussions in addition to studying normative democratic outcomes, especially in these highly polarized times.

We observe that many participants, as a form of mental self-preservation, aim to have dispassionate discussions and sometimes even preemptively disengage if they feel that they or their interlocutors are getting emotional, for fear that emotions could devolve into name-calling. They

make a distinction between being "emotional" and "rational". However, this hyper-rational, impersonal style of deliberation could have unintended consequences. Firstly, research suggests that taking emotions out of political discussions does not necessarily lead to more rational outcomes; while anger spurs aversion and leads to close-mindedness, when the emotional response is anxiety, people seek new perspectives and become open to compromise [42]. Anger, on the other hand, also increases political participation [78]. Secondly, as Young notes, "a norm of dispassionateness dismisses and devalues embodied forms of expression, emotion, and figurative expressions. People's contributions to a discussion tend to be excluded from serious consideration not because of what is said, but how it is said." [82] Clearly, this limits whose views are engaged with in cross-partisan interactions; users who are directly affected by the discussed issues likely passionately voice their opinions while those that are unaffected likely remain detached. Thus, the views of users who have the highest stakes may be less attended to. Finally, pure reasoning, with its emphasis on rationality as opposed to passion is known to be also exclusionary towards members of disadvantaged groups and individuals with less formal education as this form of communication is deliberately learned and developed [48]. Important questions around the outcomes of these conversations remain; does this kind of cross-partisan discourse contribute positively to building a deliberative democracy? In its current form, the prevailing hostility, toll on mental health, and the possible unintended consequences of participants' strategies to have deliberative discussions suggest otherwise–at least on Reddit.

5.3 Impacts of information about interlocutors: humanizing, stereotyping, judging, and attacking

In our interviews before showing the design probes, participants readily acknowledged that knowing more about others could be humanizing, allowing them to "see a little more humanity in what is otherwise a username and text" (P06). However, in current practice, participants described that they exclusively focused on the comment text and not on the author of the comment due to the fear that they may become prejudiced by viewing the author's past behavior or positions. Given that the Reddit interface does not provide the means to only see humanizing information while avoiding prejudice-inducing information, participants resolve this issue by simply not viewing user profiles entirely. This reduces opportunities to build common ground and trust. We aimed to address this issue by showing potentially humanizing information about the user through our designs.

Upon viewing the user card, as expected, participants indicated that it would alter how they participate in conversations, making them consider both the comment and the comment's author when responding. In the case of shared subreddits, many predicted that this shift could humanize the interlocutors and promote goodwill, with participants expressing that they would be more mindful of their behavior and more charitable of their interlocutors' potential transgressions. However, participants also expressed many significant concerns about other ways that the information might be used. With active subreddits, they worried that other users might judge/stereotype them negatively for participating in casual subreddits such as meme subreddits and also felt this component disclosed too much information about them. For comment highlights, female and minority participants were especially concerned about how these comments could provide more fodder to attack them. Some of these concerns can be addressed by providing users with more control over what information is shown about them on the user card.

However, more broadly, focusing attention on the user profiles, while potentially humanizing especially when users share group memberships, could have major negative implications especially for female and minority users by increasing visibility and inviting increased scrutiny on their profiles. Although participants expressed frustration over Reddit's anonymity providing a safe

harbor for trolls, they also appreciated how this anonymity allowed them to express opinions without being targeted.

5.4 Challenges and opportunities for future designs to improve cross-partisan discussions

Given the concerning feedback that we received, we decided not to proceed with building the proposed browser extension. Instead, we detail challenges in designing to improve cross-partisan discussions and opportunities we see to move forward in this space.

5.4.1 Countering the different forms of hostility in cross-partisan discussions. Partisan attacks and name-calling during cross-partisan discussions are commonplace online. Our designs aimed to minimize such occurrences by highlighting non-political group memberships to offset the effects of outgroup categorizations. However, it is important to consider other forms of hostility. For example, determined users can search through interlocutors' activity history to find material to disrupt the discussion and attack them. These concerns are particularly significant for many of our female and minority participants for whom partisan hostility often interacts with sexism and racism in cross-partisan interactions. The culture of harassment based on toxic conceptions of race, gender and sexual identities supported by Reddit's design and governance, which Massanari terms as "toxic technocultures" [46], negatively influence and exacerbate the hostility already prevalent in these political discussions. Future work on designs to improve cross-partisan discourse should attend to the multiple forms of hostility prevalent and how the designs to reduce hostility may differentially impact members of disadvantaged and marginalized social groups.

5.4.2 Countering party-stereotypes without revealing user information. Our designs aimed to cross-categorize and decategorize at an individual level by making certain user activity more visible. However, as we showed in Section 3.5.1, not all user interacting pairs have common group memberships. Further, in the previous section, we highlighted some concerns about revealing user information, and one user went so far as to say she would make throwaway accounts to disrupt that. Thus, alternate approaches to de-stereotype without calling attention to individual profiles may be more effective. For example, Alher et al. [1] found that people consistently overestimate the extent to which party supporters belong to party-stereotypical groups, sometimes by over 300% (for example, atheists for Democrats and evangelical for Republicans) and correcting these misconceptions led to significant reductions in outparty hostility. Similarly, we could surface subreddit memberships in aggregate to counter some of these extreme stereotypes. For example, by showing that only (a surprisingly low) 5.34% of Reddit users who participate in r/Conservative also participate in r/Christianity (based on 2019 Reddit comment data).

5.4.3 Intervening in cross-partisan discussions. In recent years, researchers have developed algorithms to detect when a conversation is likely to go awry to encourage either the moderators or the conversation participants to possibly take course correction measures (such as [10, 41]). However, from our interviews, participants' own attempts at de-escalating often either have little effect or cause more harm (Section 4.3). Designers aiming to intervene in individual discussions must evaluate if and when to intervene, taking into account the possible adverse effects of their interventions. We recommend that such systems allow users to choose if they want intervention in the first place so that they help facilitate and prolong discussions that the users actually want to participate in and help exit ones that they do not. However, given the low rates of success for turning around a conversation and the possibility of unintended harm, we urge designers to explore preventive measures such as improving community norms around deliberation rather than corrective measures to improve individual discussions.

5.4.4 Information Asymmetry. Our designs aimed to provide more information about interlocutors to facilitate better discussions. As outside researchers do not have access to the site, these designs are typically deployed as browser extensions or external apps. However, such a deployment would result in some users (who download the extension) having easy access to information about others, while other users may not even know that their interlocutors have access to their information. Further, even if the extension allowed users to customize or remove content on their user cards, users first would need to know that such an extension exists and download it. This will likely compound concerns that users already have about revealing more information about them. Given that cross-partisan interactions often turn into adversarial situations, one approach could be to apply an affirmative consent lens to design, centering individual agency with interactions structured around consent that is voluntary, informed, revertible, specific, and unburdensome [29]. Thus, a possible modification could be that users be able to view subreddit participation details of only other extension users who consent to information sharing. This change may necessitate a user recruitment strategy where extension users have a high likelihood of interacting with each other. To maximize the chances of such interaction, the deployment could be targeted to users participating in a particular subreddit.

6 LIMITATIONS

Our study focuses on cross-partisan discussions on Reddit only; future work on other platforms will surely improve our understanding. Given our participants' strong support for showing shared group memberships between Reddit users who are essentially strangers, we expect that showing such connections on Facebook, especially between weak ties, will have a similar impact. However, we expect that showing other active group memberships will have little impact on Facebook as users already have access to some individuating information about others in the form of a real name, profile picture and cover photo, unlike on Reddit where users typically only identify themselves using a username.

Our study is US-centric and was conducted in highly polarized times, during the lead-up to one of the most contentious US presidential elections in history. In less polarized countries/settings, partisan identities will be less salient; we speculate that these designs would have smaller effects on reducing hostility in interactions, as partisan group dynamics is unlikely to be the cause for the hostility. Alternately, approaches aimed at establishing more deliberative discussion norms through example setting [73] may be more effective as these norms might face less resistance from the hostile partisan norms that we observe today.

While in this work, we have focused on cross-partisan political discussions online, we do not contend that cross-partisan interactions are more important or should take primacy over other forms of political discourse that in some cases specifically exclude dissenting voices. As, Mansbridge et. al [43] note:

Activist interactions in social movement enclaves are often highly partisan, closed to opposing ideas, and disrespectful of opponents. Yet the intensity of interaction and even the exclusion of opposing ideas in such enclaves create the fertile, protected hothouses sometimes necessary to generate counter-hegemonic ideas. These ideas then may play powerful roles in the broader deliberative system, substantively improving an eventual democratic decision.

7 CONCLUSION

In this work, we have explored how users navigate the contentious political climate in the US to engage in cross-partisan discussions. We find that participants have different, multiple motivations for engaging in these interactions; sometimes they prefer serious deliberative discussions and other times, they look for entertainment and banter. These different motivations coupled with

the hyper-partisan environment present challenges to participants seeking to engage with "the other side". Through experience, participants have developed multiple strategies to foster good conversations. From our design probes, we observe that participants find shared non-political subreddit memberships of their interlocutors humanizing, however, sharing other details such as other group subreddit memberships and past top comments raise significant concerns around privacy and misuse.

8 ACKNOWLEDGEMENTS

We thank Libby Hemphill, Angela Schöpke-Gonzalez and Vaishnav Kameswaran for their detailed reviews and suggestions on earlier drafts of the paper. We also thank the anonymous reviewers for their incredibly useful feedback over multiple revisions which has shaped this work immensely. Ashwin Rajadesingan is supported by a Facebook Fellowship. This material is based upon work supported by the National Science Foundation under Grant No. IIS-1717688.

REFERENCES

[1] Douglas J Ahler and Gaurav Sood. 2018. The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics* 80, 3 (2018), 964–981.

[2] Kimberley Allison and Kay Bussey. 2020. Communal quirks and circlejerks: A taxonomy of processes contributing to insularity in online communities. In *Proceedings of the International AAAI Conference on Web and Social Media*.

[3] Monica Anderson and Brooke Auxier. 2020. 55% of U.S. social media users say they are 'worn out' by political posts and discussions. *Pew Research Center 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/* (2020).

[4] André Bächtiger, John S Dryzek, Jane Mansbridge, and M Warren. 2018. Deliberative Democracy. *The Oxford handbook of deliberative democracy* (2018), 1.

[5] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 830–839.

[6] Marilynn B Brewer. 1984. Beyond the contact hypothesis: Theoretical perspectives on desegregation. *Groups in contact: The psychology of desegregation* (1984), 281–302.

[7] Marilynn B Brewer. 2000. Reducing Prejudice Through Cross-Categorization: Effects. *Reducing prejudice and discrimination* (2000), 165–185.

[8] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

[9] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.

[10] Jonathan P Chang and Cristian Danescu-Niculescu-Mizil. 2019. Trouble on the Horizon: Forecasting the Derailment of Online Conversations as they Develop. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 4745–4756.

[11] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. sage.

[12] 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*. 1217–1230.

[13] Richard J Crisp and Miles Hewstone. 2007. Multiple social categorization. *Advances in experimental social psychology* 39 (2007), 163–254.

[14] Richard J Crisp, Judi Walsh, and Miles Hewstone. 2006. Crossed categorization in common ingroup contexts. *Personality and Social Psychology Bulletin* 32, 9 (2006), 1204–1218.

[15] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8.

[16] Daniel DellaPosta, Yongren Shi, and Michael Macy. 2015. Why do liberals drink lattes? *Amer. J. Sociology* (2015).

[17] James N. Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2021. (Mis-)Estimating Affective Polarization. *The Journal of Politics* 0, ja (2021), null. https://doi.org/10.1086/715603 arXiv:https://doi.org/10.1086/715603

[18] M. Duggan and A. Smith. 2016. The tone of social media discussions around politics. *Pew Research Center http://www. pewinternet. org/2016/10/25/the-tone-of-social-media-discussions-around-politics* (2016).

[19] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1175–1184.

[20] Casey Fiesler, Jialun Jiang, Joshua McCann, Kyle Frye, and Jed Brubaker. 2018. Reddit rules! characterizing an ecosystem of governance. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[21] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *interactions* 6, 1 (1999), 21–29.

[22] Bryan T Gervais. 2015. Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics* 12, 2 (2015), 167–185.

[23] 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*. 1400–1408.

[24] Jurgen Habermas and Jürgen Habermas. 1991. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press.

[25] Jeffrey T Hancock, Catalina L Toma, and Kate Fenner. 2008. I know something you don't: the use of asymmetric personal information for interpersonal advantage. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 413–416.

[26] Carolyn M Hendriks, John S Dryzek, and Christian Hunold. 2007. Turning up the heat: Partisanship in deliberative innovation. *Political studies* 55, 2 (2007), 362–383.

[27] Michael A Hogg and John C Turner. 1987. Intergroup behaviour, self-stereotyping and the salience of social categories. *British Journal of Social Psychology* 26, 4 (1987), 325–340.

[28] Leonie Huddy and Alexa Bankert. 2017. Political partisanship as a social identity. In *Oxford research encyclopedia of politics*.

[29] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark Ackerman, and Eric Gilbert. 2021. Yes: Affirmative Consent as a Theoretical Framework for Understanding and Imagining Social Platforms. (2021).

[30] Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood. 2019. The origins and consequences of affective polarization in the United States. *Annual Review of Political Science* 22 (2019), 129–146.

[31] Shanto Iyengar, Gaurav Sood, and Yphtach Lelkes. 2012. Affect, not ideologya social identity perspective on polarization. *Public opinion quarterly* 76, 3 (2012), 405–431.

[32] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. " Did You Suspect the Post Would be Removed?" Understanding User Reactions to Content Removals on Reddit. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–33.

[33] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online harassment and content moderation: The case of blocklists. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 2 (2018), 1–33.

[34] Joohan Kim and Eun Joo Kim. 2008. Theorizing dialogic deliberation: Everyday political talk as communicative action and dialogue. *Communication Theory* 18, 1 (2008), 51–70.

[35] Samara Klar, Yanna Krupnikov, and John Barry Ryan. 2018. Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship. *Public Opinion Quarterly* 82, 2 (2018), 379–390.

[36] Robert E Kraut, John M Levine, Marisol Martinez Escobar, and Amaç Herdağdelen. 2020. What Makes People Feel Close to Online Groups? The Roles of Group Attributes and Group Types. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 382–392.

[37] 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*. ACM.

[38] 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*. ACM, 1559–1568.

[39] Alex Leavitt. 2015. " This is a Throwaway Account" Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 317–327.

[40] Matthew S Levendusky. 2020. Our common bonds: Using what Americans share to help bridge the partisan divide. *Unpublished Manuscript, University of Pennsylvania* (2020).

[41] P Liu, J Guberman, L Hemphill, and A Culotta. 2018. Forecasting the presence and intensity of hostility on Instagram using linguistic and social features. In *Proceedings of the 12th International Conference on Web and Social Media*.

[42] Michael MacKuen, Jennifer Wolak, Luke Keele, and George E Marcus. 2010. Civic engagements: Resolute partisanship or reflective deliberation. *American Journal of Political Science* 54, 2 (2010), 440–458.

[43] Jane Mansbridge, James Bohman, Simone Chambers, Thomas Christiano, Archon Fung, John Parkinson, Dennis F. Thompson, and Mark E. Warren. 2012. A systemic approach to deliberative democracy. *Deliberative systems: Deliberative democracy at the large scale* (2012), 1–26.

[44] Alice E Marwick and Danah Boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133.

[45] Lilliana Mason. 2016. A cross-cutting calm: How social sorting drives affective polarization. *Public Opinion Quarterly* 80, S1 (2016), 351–377.

[46] Adrienne Massanari. 2017. # Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society* 19, 3 (2017), 329–346.

[47] Adrienne Lynne Massanari. 2015. Participatory culture, community, and play. *Learning from* (2015).

[48] Chantal Mouffe. 2000. Politics and Passions. *Ethical Perspectives* 7, 2-3 (2000), 146–150.

[49] Ashley Muddiman and Natalie Jomini Stroud. 2017. News values, cognitive biases, and partisan incivility in comment sections. *Journal of communication* 67, 4 (2017), 586–609.

[50] Sean A Munson, Stephanie Y Lee, and Paul Resnick. 2013. Encouraging reading of diverse political viewpoints with a browser widget. In *Seventh international aaai conference on weblogs and social media*.

[51] Sean A Munson and Paul Resnick. 2010. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1457–1466.

[52] Matti Nelimarkka, Jean Philippe Rancy, Jennifer Grygiel, and Bryan Semaan. 2019. (Re) Design to Mitigate Political Polarization: Reflecting Habermas' ideal communication space in the United States of America and Finland. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[53] Matti Nelimarkka, Jean Philippe Rancy, Jennifer Grygiel, and Bryan Semaan. 2019. (Re) Design to Mitigate Political Polarization: Reflecting Habermas' ideal communication space in the United States of America and Finland. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 141.

[54] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you.* Penguin UK.

[55] Thomas F Pettigrew. 1998. Intergroup contact theory. *Annual review of psychology* 49, 1 (1998), 65–85.

[56] Tom Postmes and Nancy Baym. 2005. Intergroup dimensions of the Internet. *Intergroup communication: Multiple perspectives* 2 (2005), 213–240.

[57] Tom Postmes, Russell Spears, and Martin Lea. 2002. Intergroup differentiation in computer-mediated communication: Effects of depersonalization. *Group Dynamics: Theory, Research, and Practice* 6, 1 (2002), 3.

[58] Stephen A Rains, Kate Kenski, Kevin Coe, and Jake Harwood. 2017. Incivility and political identity on the Internet: Intergroup factors as predictors of incivility in discussions of news online. *Journal of Computer-Mediated Communication* 22, 4 (2017), 163–178.

[59] Ashwin Rajadesingan, Ceren Budak, and Paul Resnick. 2021. Political Discussion is Abundant in Non-political Subreddits (and Less Toxic). In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 525–536.

[60] Ashwin Rajadesingan, Paul Resnick, and Ceren Budak. 2020. Quick, Community-Specific Learning: How Distinctive Toxicity Norms Are Maintained in Political Subreddits. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 557–568.

[61] Stephen D. Reicher, Russell Spears, and Tom Postmes. 1995. A social identity model of deindividuation phenomena. *European review of social psychology* 6, 1 (1995), 161–198.

[62] Joseph Seering, Tianmi Fang, Luca Damasco, Mianhong'Cherie' Chen, Likang Sun, and Geoff Kaufman. 2019. Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 606.

[63] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping pro and anti-social behavior on twitch through moderation and example-setting. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 111–125.

[64] Bryan Semaan, Heather Faucett, Scott P. Robertson, Misa Maruyama, and Sara Douglas. 2015. Designing Political Deliberation Environments to Support Interactions in the Public Sphere *(CHI '15)*. ACM, 3167–3176.

[65] 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*. 1409–1421.

[66] 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*. ACM, 1409–1421.

[67] Jaime E Settle. 2018. *Frenemies: How social media polarizes America.* Cambridge University Press.

[68] Richard M Shafranek. 2019. Political considerations in nonpolitical decisions: a conjoint analysis of roommate choice. *Political Behavior* (2019), 1–30.

[69] Dhavan V Shah, Jaeho Cho, William P Eveland Jr, and Nojin Kwak. 2005. Information and expression in a digital age: Modeling Internet effects on civic participation. *Communication research* 32, 5 (2005), 531–565.

[70] Frank M Shipman and Catherine C Marshall. 1999. Formality considered harmful: Experiences, emerging themes, and directions on the use of formal representations in interactive systems. *Computer Supported Cooperative Work (CSCW)* 8, 4 (1999), 333–352.
[71] Simon Buckingham Shum et al. 2008. Cohere: Towards web 2.0 argumentation. *COMMA* 8 (2008), 97–108.
[72] Elizabeth Suhay, Emily Bello-Pardo, and Brianna Maurer. 2018. The polarizing effects of online partisan criticism: Evidence from two experiments. *The International Journal of Press/Politics* 23, 1 (2018), 95–115.
[73] 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*. 3401–3410.
[74] Martin Tanis and Tom Postmes. 2005. A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology* 35, 3 (2005), 413–424.
[75] W Ben Towne and James D Herbsleb. 2012. Design considerations for online deliberation systems. *Journal of Information Technology & Politics* 9, 1 (2012), 97–115.
[76] John C. Turner, Rupert J. Brown, and Henri Tajfel. 1979. Social comparison and group interest in ingroup favouritism. *European journal of social psychology* 9, 2 (1979), 187–204.
[77] John C Turner, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell. 1987. *Rediscovering the social group: A self-categorization theory.* Basil Blackwell.
[78] Nicholas A Valentino, Ted Brader, Eric W Groenendyk, Krysha Gregorowicz, and Vincent L Hutchings. 2011. Election night's alright for fighting: The role of emotions in political participation. *The Journal of Politics* 73, 1 (2011), 156–170.
[79] Sai Wang. 2020. The Influence of Anonymity and Incivility on Perceptions of User Comments on News Websites. *Mass Communication and Society* 23, 6 (2020), 912–936.
[80] Magdalena Wojcieszak and Benjamin R Warner. 2020. Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact. *Political Communication* (2020), 1–23.
[81] Michael Yeomans, Julia Minson, Hanne Collins, Frances Chen, and Francesca Gino. 2020. Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes* (2020).
[82] Iris Marion Young. 2002. *Inclusion and democracy.* Oxford University press on demand.
[83] Elena Zheleva and Lise Getoor. 2009. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th international conference on World wide web*. 531–540.

Received October 2020; revised April 2021; accepted July 2021