Capturing Diverse and Precise Reactions to a Comment with User-Generated Labels

Eun-Young Ko*
School of Computing, KAIST
Daejeon, Republic of Korea
eunyoungko@kaist.ac.kr

Eunseo Choi*
MIT
Cambridge, MA, USA
choie@mit.edu

Jeong-woo Jang
School of Humanities and Social Sciences, KAIST
Daejeon, Republic of Korea
jangjw29@kaist.ac.kr

Juho Kim
School of Computing, KAIST
Daejeon, Republic of Korea
juhokim@kaist.ac.kr

ABSTRACT

Simple up/downvotes, arguably the most widely used reaction design across social media platforms, allow users to efficiently express their opinions and quickly evaluate others' opinions from aggregated votes. However, such design forces users to project their diverse opinions onto dichotomized reactions and provides limited information to readers on why a comment was up/downvoted. We explore user-generated labels (UGLs) as an alternative reaction design to capture the rich context of user reactions to comments. We conducted a between-subjects study with 218 participants to understand how people use and are influenced by UGLs compared to up/downvotes. Specifically, we examine how UGLs affect users' ability to express and perceive diverse opinions. Participants generated 234 unique labels on diverse aspects of a comment. Leaving more reactions than participants in the up/downvotes condition, participants reported that the ability to express their opinions improved with UGLs. UGLs also enabled participants to better understand the multifacetedness of public evaluation of a comment.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; **Social media**.

KEYWORDS

computer-mediated communication, online discussion, social computing, social media, reaction buttons

ACM Reference Format:
Eun-Young Ko, Eunseo Choi, Jeong-woo Jang, and Juho Kim. 2022. Capturing Diverse and Precise Reactions to a Comment with User-Generated Labels. In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3485447.3512243

*Both authors led this work together.

WWW '22, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9096-5/22/04…$15.00
https://doi.org/10.1145/3485447.3512243

1 INTRODUCTION

Online comment sections open up public discourse and facilitate interactions among users, including original commenters, reactors, and viewers. As comment sections play a vital role of encouraging engagement from users, one can easily find them across the web in many social media sites. Within these comment sections, users' comments often invite subsequent comments or reactions like up/downvotes. Whereas reply comments allow informationally rich input, reactions provide easier ways to express one's opinion. Once accumulated, such reactions influence the way users perceive and interact with the comments [36, 44], their perception of the public opinion [20, 27, 28, 31, 48], and their views of the relevant topic [24, 31, 41, 45].

Whereas up/down votes are the most widely used form of reaction, it is questionable whether they provide sufficient context to readers on why or how people are responding to comments. In our formative study, which aimed to understand how people use and interpret up/downvotes, we found that people's considerations for diverse aspects of a comment were projected onto up/downvotes. Our interviewees noted that they were unable to precisely express what they felt, especially when they have mixed or moderate opinions, and this discouraged them from leaving reactions to comments. However, when they were asked to share their interpretation of aggregated votes, most interviewees simplistically interpreted up/downvotes as users' (dis)agreement to a comment.

In this work, we explore the potential of having users cooperatively generate labels that capture different aspects of a comment that people focus on when they react to a comment. We suggest user-generated labels (UGLs) (Figure 1) as an alternative reaction design that captures rich context of user reactions. For each comment, users can create their own UGLs or add votes to UGLs generated by other users to efficiently indicate what aspect of the comment deserves an up/downvote.

We conducted a between-subjects study with 218 participants to explore the effect of UGLs in capturing diverse and precise reactions and users' understanding of multifacetedness of reactions to a comment. We compared the baseline condition, in which participants can leave up/downvotes and reply to a comment, and the UGL condition, in which participants can generate UGLs, vote on UGLs, and reply to a comment. In the UGL condition, participants generated 234 unique labels regarding the degree of agreement, the strength of the argument, the style of the comment, judgments on the commenter, and feelings or beliefs related to the topic. With UGLs, participants felt that their reactions were more precise and unique, and left more reactions. Moreover, participants better understood the multifacetedness of others' reason for reacting to a comment with UGLs compared to the baseline condition.

The contributions of the paper are as follows:

- User-generated labels (UGLs), a reaction design that captures diverse and precise reactions to a comment.
- Observations from a formative study and a user study that provide insights into dimensions users consider when they react to comments.
- Findings from a user study that show (1) UGLs enhance users' ability to express and interpret others' reactions, and (2) users leave more reactions and better understand the multifacetedness of opinion towards a comment.

2 BACKGROUND AND RELATED WORK

In this section, we review previous work that explored the effect of diverse reaction designs. Then we discuss previous work on leveraging the intrinsic motivation of users in social systems. Lastly, we discuss the affective polarization and its relation to the perceived diversity in public opinion.

2.1 Choice of Reaction Buttons

While up/downvotes are most commonly used, there are different types of reactions available across various platforms. These include like/dislike, emojis, recommend, and the heart button. However, binary buttons like up/downvotes may vary in their meanings across contexts, sites, and users [43]. Inconsistency across users' intention for choosing from pre-set reactions causes lack of clarity. Emojis are often used ironically; individuals may also use them to signal different levels of agreement [31]. As such, current reaction modalities do not effectively present these multidimensional messages, limiting users' abilities to infer information behind the reactions.

Prior work suggested different preset categories. Sumner et al. [2018] suggested Interesting, Amused, and Love buttons as possible additions to a like button for users to be more specific about their positive connection with a comment's content [40]. Nevertheless, these categories still cannot cover evaluation of a comment with mixed sentiments. Based on the Stereotype Content Model, Stroud et al. [2017] proposed the use of "Respect" over "Like" and "Recommend" as an alternative reaction button to reduce users' hostility towards comments with opposite political attitudes [38]. However, "Respect" only covers a relational connection with the comment; it forgoes the evaluative functionality of an upvote.

Evidently, reactors' ability to express is hindered in current social media comment sections as available reactions fail to coherently communicate to other users as to why reactors upvoted a comment. Without placing too much cognitive load on their users, UGLs aim to relay crowds' underlying sentiment and evaluation of a comment in more detail across users.

2.2 Designing Incentives for User Participation

User comments and reactions are valuable information that people read and interpret to understand public opinion and develop their thoughts. However, only a small portion of users leave comments or reactions and most users prefer to lurk, observing other people's thoughts but not sharing their own [32, 39]. Although lurking is a natural and valuable activity that makes the shared opinions heard [29], user participation in adding their own comments and reactions is essential to maximize this public good [6].

Previous studies have explored diverse social-psychological incentives to promote users' participation. Researchers found a positive relationship between individual engagement online and social factors such as expectation of social approval [6, 13], social transparency [14], and reciprocity [5]. Other studies have shown that intrinsic factors such as enjoyment [21, 26], self-expressiveness [34], and perceived impact [2] and uniqueness [4, 9, 22] of individual contribution increase user participation.

In this paper, we explore an alternative reaction design that can capture users' detailed opinions on a comment. As describing one's thoughts in detail requires more cognitive effort from users, it should be designed in a way that triggers users' intrinsic motivation to engage. Reacting through UGLs enables users to create labels in their own words and positively affects the level of perceived uniqueness of reaction and self-expressiveness.

2.3 Affective Polarization and Understanding the Diversity in Public Opinion

Prior work has shown that people overestimate the difference between "the other party" and themselves [19] and underestimate the level of agreement with those who take the opposite stance on an issue [10, 11, 35]. By overestimating disagreement with members from the opposing group, people often become less tolerant toward the outgroup, which lowers their willingness to socialize and have a constructive discussion with the opposite party [10, 35].

The idea of increasing exposure to different points of view is a common intervention suggested by previous studies to decrease polarization. While exposure to diversity is not sufficient by itself to eliminate polarization, many researchers have seen its effect on increasing understanding and decreasing dislike between inter-group members [1, 30]. Our work revisits the effect of exposure to diversity. Whereas most work [1, 3, 17, 18] focuses on the diversity of opinions towards a topic, we focus specifically on the diversity of reactions towards a comment and its effect on reducing affective polarization among users. User-generated descriptive labels have two aims: improve users' understanding of which aspect a comment is valuable and organically set up a space for intergroup members to share their diverse judgment about a comment. We explore whether UGLs affect users' hostility and naive realism, the tendency to consider others' views as homogeneous and incorrect [33]. Specifically, we investigate how UGLs affect users' understanding of multifaceted public evaluation of a comment and tolerance to the opinions that do not align with theirs.

3 FORMATIVE STUDY

To understand how users leave and interpret reactions to comments through dichotomous reaction buttons, we conducted a series of observations and semi-structured interviews. We recruited 10 participants (6 undergrads, 4 grad students) from an online community of a technical university in South Korea.

3.1 Task and Procedure

We implemented a toy comment system where participants can read a news story and user comments, reply and react to user comments, or leave their own comments. We chose three news stories on controversial issues in Korea at the time of study (legalizing abortion, distributing subsidies for COVID-19, and location-tracking to combat COVID-19) from Naver[1]. For each story, we chose ten comments (5 supporting and 5 opposing) from a pool of comments with the most number of votes (up/downvotes).

In each session, participants were asked to think aloud as they read each news story and relevant user comments. During this process, they were able to add comments, reply to comments, or leave up and downvotes. To observe participants' original reaction to a comment without social influence, we hid previously accumulated up/down votes on each comment. After the think-aloud session, we conducted a semi-structured interview and asked participants to describe their positions on the issue and give reasons for their reactions or inactions to each of the comments. At the end of each interview, we revealed the aggregated reactions to the same set of comments to our participants who were then asked to share their interpretation of the reaction statistics along with their perception of the general public opinion.

3.2 Observations

We observed that current reaction buttons do not fully capture nor represent users' evaluative judgments about the comment. We present the main findings in more detail below.

3.2.1 A simple up/downvote does not capture nuances and diverse dimensions of people's actual reactions to comments. Interviewees had different rationales for using the same reaction button. For example, P3 said he downvotes to signal disagreement with the claim made in a comment. P2 said he downvotes when a rationale provided in a comment is not valid. P1 gave an upvote for effort, while P5 upvoted a comment for humor. All of these diverse evaluation aspects were projected onto either an upvote or downvote.

Also, there were individual differences in people's thresholds in voting a comment. When interviewees partially agreed or disagreed with a comment, some people projected that opinion into an up/downvote while others did not. Some interviewees did not leave reactions when they have a mixed evaluation of a comment (e.g., disagree but logical). It is possible to express these mixed evaluations through a reply comment, but participants said that they reply to comments only when they feel the strong urge to say something. Compared to reactions, we can assume that replying to comments requires much more effort from the user and is only utilized when users have a high willingness to express themselves.

3.2.2 Willingness to react decreased when participants cannot precisely express their opinions or make a meaningful contribution. We noticed a number of factors that decreased people's willingness to react to a comment. For some interviewees, moderate or mixed evaluation of a comment did not pass their thresholds to vote a comment. An interviewee explicitly said he would not react if it is difficult to accurately articulate his opinion.

We also observed that people's willingness to react decreased when their perceived magnitude of contribution through participation is low. This was especially true when up/downvotes accumulated to sufficient amounts. Some interviewees said that the marginal impact their inputs could have on others mattered in their decision to react to a comment. This was especially true for comments with many votes, more so when votes were skewed to one side. They felt that even if they downvoted a comment to express their disagreement, it would not change the majority's positive perception of a comment that already has many upvotes.

3.2.3 Users assume that up/downvotes represent reactors' stance about the topic in relation to the comment. Interviewees placed importance on different aspects of a comment, yet such variance across people's value judgments was indistinguishable when aggregated into up/downvotes. When interviewees were asked to share their thoughts on why some comments got up/downvotes, 7 out of 10 participants said that people who agreed or disagreed with the comment left up/downvotes. Moreover, despite the complexity of people's reasons for reacting to comments, none of our interviewees considered the possibility of mixed evaluation to a comment. These observations imply that up/downvotes fail to deliver reactors' nuanced opinions to readers. We also saw some users expressing greater contempt towards the outgroup as a result. When an interviewee saw many likes attached to contentious comments contradicting his opinion, he immediately called out, "Anyone who liked this comment is probably a misogynist."

4 USER-GENERATED LABELS

We introduce the concept of user-generated labels (UGLs) that enable users to create text-based reactions. Users can make UGLs to describe their thoughts on a comment in short text and vote on UGLs created by others. We suggest UGLs as an alternative reaction design that captures diverse and precise reactions to a comment in a light-weighted interaction. In this section, we explain how users can interact with UGLs and then discuss how UGLs solve problems identified in our formative study.

4.1 Reaction through UGLs

To each comment, a user can add descriptive UGLs or click on existing ones. As in Figure 1, UGLs are generated and displayed right below the comment.

Creating UGLs Users can add their own UGL(s) by clicking an add button (button with + mark in Figure 1) and writing a short text. Users self-classify each of their UGL as either a positive or negative reaction. To guide users to make short and descriptive labels, we impose character limits (20 characters) and do not allow spaces (CamelCase) for each UGL. Generated UGLs are shared with other users and also serve as a voting button. To prevent users from making duplicated labels, users' input is matched with existing UGLs through autocomplete when applicable.

Reading and voting on UGLs Below each comment, readers can read positive evaluations of a comment from the blue-colored pane on the left and negative evaluations from the red-colored pane on the right. We included this dichotomy to organize generated UGLs such that it is easier for users to read and vote on UGLs with specific sentiment (positive or negative). Indicated sentiment

[1] Naver is the largest news platform in South Korea with an active comment section. https://news.naver.com

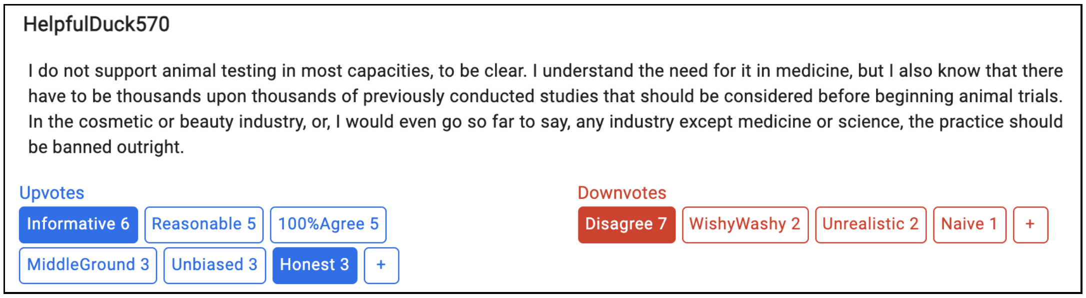


Figure 1: User-generated labels (UGLs) allows users to express their reactions to a comment by either creating their own text-based UGL or by voting on UGLs generated by other users.

also helps other users to better understand UGLs that are not self-explanatory (e.g., SlipperySlope). Users can click on one or more UGLs to add their votes. For example, it is possible for the user to vote for two positive UGLs and one negative UGL on a comment as in Figure 1. Positive and negative evaluations of a comment from reactors are organized and ordered by the number of votes.

Managing toxic or irrelevant UGLs In the pilot study we conducted during the iterative design process, we found that some users generate irrelevant or offensive UGLs. Since UGLs serve as voting options for users and are highly visible right below the comments, such inappropriate UGLs can harm other users' experience. To limit the abuse of UGLs and control for incivility, users can right-click and flag a UGL if they find it inappropriate. When a user flags UGLs, the system asks the user to choose the reason for flagging among irrelevant, insulting, or other (explain) options. UGLs that receive more than three flags are hidden from the system, whereas the threshold can be adaptively determined for specific discussion platforms considering the number of users and the level of civility.

4.2 Rationale for the Design of UGLs

UGLs capture and present diverse and dynamic reactions to a comment. In our formative study, we found that people consider different dimensions of a comment and have their own ideas of what decreases or increases the value of a comment. UGLs enable dynamism and diversity that reflects such variety of people's viewpoints about a comment as well as their nuanced differences. Using text-based UGLs, users can precisely express and present their thoughts on a comment to other users with brevity. When a user has mixed evaluation on a comment, they can add multiple UGLs on each pane. Users can also specify their degree of (dis)agreement with the comment like 'StronglyAgree' or 'Acceptable'.

UGLs increase the perceived contribution of each reaction to a comment. The perceived level of contribution plays a key role when users decide to engage with comments and leave their opinions or evaluations. Enabling reactors to generate their own labels to comments augments their abilities to efficiently contribute, share, and exchange evaluative judgments about others' comments in an online discussion space. Users' intrinsic interests in receiving any degree of social approval can potentially be fulfilled by allowing their UGLs to be upvoted by other users. The more expressive nature of UGLs can increase users' ability to "do something about" the perceived media effect [8, 25]; as users observe more people expressing their evaluative opinions about a comment and exerting greater influence on others through UGLs, users' level of engagement can also increase.

UGLs save the effort required to express one's reaction to a comment. While the trade-off between the ability to capture diverse and precise reactions and the ease of engaging is inevitable, UGLs mitigate this trade off by allowing lightweight interactions to express one's opinion. While creating new UGLs requires a little more effort, voting on UGLs provides an easy way to express one's precise reaction to comments.

5 EVALUATION

To understand how users collectively utilize UGLs and how having UGLs affects users' experience, we conducted a between-subjects study with 218 participants. To better understand the effect of UGLs, we compared UGLs with up/downvotes. We investigate the effectiveness of UGLs with the following research questions:

RQ1: *How well do UGLs capture opinions towards comments?*

RQ2: *How does having UGLs affect user's experience in evaluating comments?*

Then we expand our investigation by exploring the secondary effect of UGLs on users' perception of the public opinion and outgroups:

RQ3: *Do UGLs allow users to better understand the multifacetedness of public evaluation of a comment?*

RQ4: *How do UGLs affect users' tolerance to the opinions that do not align with theirs?*

5.1 Study Design

We conducted a between-subjects study with two conditions: Binary and UGL. In both conditions, participants were given a topic statement and six initial comments (3 supporting, 3 opposing the topic statement). Participants in the Binary condition were able to react to the comments through up/downvotes. In the UGL condition, participants were able to react to the comments through UGLs. In both conditions, participants were able to add a comment or leave a reply to existing comments. Participants could also see others' reactions and reply comments.

We ran a study with four discussion topics: 1) Banning capital punishment 2) Banning affirmative action in hiring practice 3)

Banning animal testing 4) Regulating tech companies' use of consumer data. These topics were chosen based on the pre-study survey that we ran on Amazon Mechanical Turk (AMT). We presented 13 different discussion topics and asked participants to share their stance, relevance, and importance (in a 7-point Likert scale) and their opinions (in a short paragraph) on each topic. We collected responses from 30 respondents and chose four topics with evenly distributed opinions. For each topic, we selected 6 open-ended responses (3 supporting, 3 opposing) and used them as initial comments so that users who joined the system in the early stage can still have some comments to read and react to.

5.1.1 Participants. We recruited 218 participants from AMT and randomly assigned them into one of eight groups (4 topics, 2 experimental conditions). The number of participants in each group varied from 24 to 31. We had 109 participants on the Binary and UGL condition each. Participants were paid $4 for their participation in a one-time, 30-minutes long session. The average age of participants was 40.7 (SD: 11.2). We had 139 participants with university education, and 21 of them had post-graduate education.

5.1.2 Tasks and Procedure. Before the main activity, participants were asked to answer the pre-survey that asked how often they participate in an online discussion (reading, writing, or reacting to a comment). In addition to their stance, relevance, importance, and willingness to express on the given topic issue, we also asked questions on participants' level of tolerance for people with opposite stance in 7-point Likert scale questions.

At the start of the main activity, we explained to participants that they can comment, reply, or react to a comment as in a common discussion platform. Participants then entered a system where they were presented with six initial comments. They were also able to see earlier participants' comments, replies, and reactions. To replicate real-life comment sections and examine how users' behavior changes as reactions accumulate over time, participants' reactions were saved and were shown to subsequent participants. While participants were told that they could read and participate in the online discussion (comment, reply, or react) as long as they wanted, they had to wait at least two minutes until they could move on to the post-survey link. We designed the time constraint to get participants engaged in the discussion, not necessarily forcing them leave comments, replies, or reactions.

In the post-survey, we asked questions on participants' experience in reacting to a comment or seeing other participants' reactions. Then participants were asked to explain what other people's reasons for up/downvoting a comment would be. We also asked questions on usability, perceived multifacetedness of users' opinion on a comment, and tolerance to people with opposite stance.

5.2 Measures

We used multiple measures that operationalize variables of interest with regards to each RQ. Below we summarize measures that we used to answer each RQ.

To answer **RQ1** we analyzed (1) the diversity in UGLs, (2) the number of reactions, and (3) the number of comments that each participant reacted to. To measure the diversity in UGLs, we first generated the list of unique UGLs by manually merging labels with the same meaning (e.g., 'Misinformation' and 'WrongInformation') into one. Then we categorized unique UGLs based on the aspects each UGL focuses on. To establish the set of categories, one researcher conducted an open coding on half of UGLs and another researcher reviewed the categories. Table 1 shows the established set of categories. Then the two researchers independently coded every UGLs, and they compared and discussed to resolve conflicts. The inter-rater reliability was 0.71 (Cohen's κ, with SE: 0.04). To see how UGLs affect the number of reactions, we compared the number of reactions left on the six initial comments, which were presented to all participants in both conditions. Lastly, we counted the number of comments that each participant reacted to among the six initial comments.

Regarding **RQ2**, we distributed 7-point Likert scale questions on their experience in leaving reaction and seeing others' reactions. For both conditions, we asked about the perceived precision of reaction and perceived uniqueness of contribution that they make through reaction. In the UGL condition, we distinguished the experience of generating a UGL from voting on others' UGLs. We also asked the perceived level of understanding of others' reactions and whether their evaluation of comments was affected by reactions from others. We then asked how mentally and physically demanding it was to react with UGLs and up/downvotes.

To see if UGLs help users better understand the multifacetedness of public evaluation (**RQ3**), we analyzed the open-ended response in which participants explained other people's reasons for up/downvoting. In our analysis, we measured the number of reasons that participants can think of as reasons behind up/downvotes and the proportion of participants who mentioned reasons other than a simple agreement. Two external raters counted the number of reasons together for the first 5% of the data, and then each external rater analyzed each half of the data separately. The same two external raters coded each response individually (first 5% of the data was coded together to build a consensus among raters), and the inter-rater reliability was 0.76 (Cohen's κ, with SE: 0.03). We also measured how they were able to see users' diverse reasons for up/downvoting a comment on a 7-point Likert scale.

Regarding **RQ4**, we asked 7-point Likert scale questions on participants' tolerance towards opinions that do not align with theirs in the pre- and post-survey. Specifically, we asked their willingness to listen to, learn from, accept the opinion of commenters with opposing stances, and join a face-to-face conversation with them [12, 16]. We also asked participants' tolerance for reactors who have sentiment towards comments that conflict with their own. We asked how much they are willing to listen others' reason for up/down voting a comment and how likely they would find these reasons justifiable. We also measured the number of positive reactions each participant left to comments with opposite stance.

5.3 Result

Overall, participants showed a moderate level of topic relevance (M: 4.55/7.00, SD: 1.94), importance (M: 4.69, SD: 1.79), and willingness to express their opinions in an online space (M: 4.05, SD: 2.06). Participants had the highest topic relevance for the consumer data topic (M: 5.50, SD:1.50) and lowest for the capital punishment topic

Table 1: The categories of UGLs with descriptions, examples, the numbers of UGLs generated, and the number of votes.

Category	Description	Examples	# UGLs generated	# Votes (% of total)
General	General labels with positive or negative sentiment	Superb, Excellent, Nah	10	52 (3.4%)
Agreement	Level of agreement or acceptance on a comment	Agree, 50%Agreed, IAcceptThis, KindaDisagee	111	517 (34.2%)
Argument	Strength of weakness of argument made in a comment	Logical, Pragmatic, Misinformation, RashAssuption	172	693 (45.8%)
Style	Writing styles or tone of a comment	TooEmotional, Vague, Forthright, HardtoRead	29	68 (4.5%)
Commenter	Judgement or impression on a commenter	WillingToListenm, Honest, Dogmatic, Compassionate	28	70 (4.6%)
Feeling	What reactors emotionally feel from a comment	Hopeful, Confused, Obnoxious	7	8 (0.5%)
Belief	Topic-specific belief or opinion	MutualBenefit, Equality,TooMuchTaxMoney	37	105 (6.9%)

(M: 3.60, SD: 2.20). However, there was no difference between Binary and UGL conditions for all four topics.

There were 109 participants for each condition. Participants generated a total of 54, 56 comments and 224, 265 replies in Binary and UGL conditions, respectively. Participants in the Binary condition made 1,027 up/down votes while UGL participants generated 394 labels and made 1,630 votes (including those votes on one's own UGLs). Numbers for each condition are reported in Appendix A.

There were 13 flags in total on the 11 UGLs. Eight flags were on the irrelevance of UGLS (e.g., TooMuchTaxMoney), and four were marked as insulting (e.g., Dumb). There were no UGLs hidden for getting three or more flags.

5.3.1 ***RQ1**: How well do UGLs capture opinions towards comments?*

Diversity in generated UGLs UGL participants generated 394 UGLs and there were 234 unique UGLs (109 positive, 125 negative). Among 394 UGLs, 'Agree' was the most frequently generated UGLs (34 times, 170 votes), followed by 'Disagree' (29 times, 96 votes) and 'Logical' (24 times, 92 votes).

Participants generated UGLs on their degree of agreement with the comment, strength of the argument, style of the comment, judgments on the commenter, and feelings or beliefs related to the topic. Table 1 shows the established categories of UGLs with description, examples, and the number of each case, and its vote counts.

Number of reactions For the six initial comments, on average, UGL participants generated 1.44 (SD: 1.90) UGLs per person and clicked on 5.51 (SD: 3.05) UGLs made by others. On the other hand, participants in the Binary condition made 4.38 (SD:1.57) votes. The difference between the number of up/down votes and the number of votes on UGLs was significant (Mann-Whitney (MW) test, Z=3.05 with p<0.005). Figure 2 illustrates the average number of up/downvotes (Binary) in comparison with UGLs (UGL) on six initial comments for each topic. There was no interaction effect between condition and topic.

However, the number of reactions in UGL condition highly depends on the number of UGLs generated at the moment of reaction. Figure 3 shows how the cumulative average number of up/downvotes (Binary), UGLs generated, and votes on UGLs (UGL) (on six initial comments) change with the accumulated number of participants for almost all topics. In the early stage when there are only a small number of UGLs that participants can vote on, the number of reactions in the Binary condition is higher than the UGL condition. Once a certain number of UGLs accumulated, UGL participants begins to leave more reactions than Binary participants. The exception was the consumer data topic in which the first UGL participants created 10 UGLs, leaving subsequent participants with more options of UGLs to vote on even in the early stage.

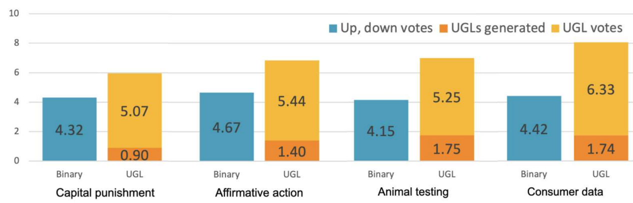


Figure 2: The average number of up/downvotes (Binary) and generated/voted UGLs (UGL) on the six initial comments.

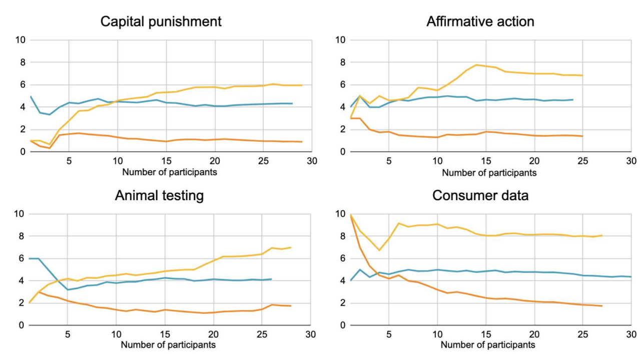


Figure 3: The cumulative average number of up/down votes (Binary), created UGLs (UGL), and votes on UGLs (UGL) by the accumulated number of participants for each topic.

Number of comments that each participant reacted to UGL participants reacted to more comments (among six initial comments) than Binary participants. UGL participants reacted to 5.12 (SD: 1.40) comments while Binary participants reacted to 4.39 (SD: 1.57) comments on average. The difference was statistically significant (MW test, Z=3.81 with p<0.0005).

Mixed evaluation captured in UGLs Out of 109 participants in the UGL condition, 14 participants left both positive and negative reactions to a comment. For example, in response to the comment "Affirmative action is necessary for a country as racist as ours is," one user labeled "Agree" and "NotThorough", which are two labels of the opposite sentiment. These 14 participants left mixed evaluation to 21 comments total.

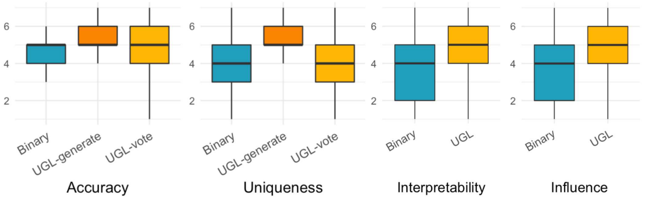


Figure 4: Perceived accuracy and uniqueness of own reaction and interpretability and influence of others' reactions, for each reaction type.

5.3.2 ***RQ2**: How does having UGLs affect users' experience in evaluating comments?* In terms of the perceived accuracy and uniqueness, participants found UGLs more satisfactory. When asked how accurately they could express their thoughts (i.e., perceived accuracy), UGL participants scored 5.24 (SD:1.35) for generating and 5.02 (SD:1.48) for voting on UGLs, while Binary participants scored 4.56 (SD:1.60). The pairwise difference between the three types of reaction was all statistically significant (MW test, Z=3.27 with p<0.001 for generating UGL vs. Binary, Z=2.01 with p<0.05 for voting on UGLs vs. Binary. Wilcoxon signed-rank (WS) test for generating UGLs vs. voting on UGLs with W=622 with p<0.05).

Participants felt that they made more unique contributions when generating UGLs (M:5.24, SD:1.40) than voting on UGLs (M:4.13, SD:1.72) or casting up/downvotes (M:4.00, SD:1.77). The difference between generating UGLs and voting on UGLs or up/downvoting was statistically significant (MW test with Z=5.36 with p<0.0001 for generating UGL vs. Binary, WS test with W=278.5 with p<0.0001 for generating UGLs vs. voting on UGLs). One UGL participant noted, *"I like that text based reactions allow me to completely show my opinion rather than just going along or against others."*. There was no difference between voting on UGLs and up/downvoting.

When asked how well they could interpret other users' opinions about a comment, UGL participants scored higher (M:4.7, SD:1.5) than Binary participants (M:3.8, SD:1.70) (MW test, Z=3.75 with p<0.0005). Moreover, when asked whether others' reactions influenced their evaluations of a comment, UGL participants reported a higher score (M:4.56, SD:1.61) than Binary participants (M:3.72, SD:1.70) (MW test, Z=3.58 with p<0.0005). *"I like seeing other's perspective and thinking about them to see if I am wrong in my thinking."* said one participant in UGL condition.

Participants reported higher mental and physical demand for generating UGLs than voting on UGLs or up/downvoting. When asked about mental and physical load needed for each type of reaction, UGL participants rated 3.42 (SD:1.73) and 2.46 (SD: 1.74) for generating UGLs, and 2.10 (SD:1.58) and 1.91 (SD:1.66) for voting on UGLs. Binary participants rated 2.09 (SD:1.64) and 1.74 (SD:1.49) for mental and physical demand, respectively. For both questions, the difference between generating UGLs and voting on UGLs (WS test, W=97 and W=187 with p<0.0001, for mental and physical) or up/downvoting (MW test, Z=6.10 and Z=4.07 with p<0.0001, for mental and physical) was statistically significant while there was no difference between voting on UGLs and up/downvoting. However, many participants described UGLs as an easy way to precisely express one's thought. *"Text based reactions seem like a very simple way to still get what you want to say about the comment. ··· I love it, a one worded comment that can still be taken as a vote."*

5.3.3 ***RQ3**: Do UGLs allow users to better understand the multi-facetedness of public evaluation of a comment?* When asked to list possible reasons behind up/downvotes, UGL participants listed 1.51 (SD:0.71) reasons for upvotes and 1.71 (SD:1.00) reasons for downvotes. Conversely, Binary participants listed 1.30 (SD:0.58) reasons for upvotes and 1.35 (SD:0.64) reasons for downvotes. The differences was significant (MW test, Z=-2.04 with p<0.01 for upvotes and Z=-2.34 with p<0.05 for downvotes).

UGL participants better understood the multifaceted aspects of up/downvoting, as 48.6% and 51.3% of them mentioned reasons behind upvotes and downvotes other than simple agreement or disagreement with a parent comment. On the other hand, only 29.4% and 33.9% of Binary participants came up with reasons behind up or downvotes other than agreement or disagreement. The differences were significant (χ^2 test, χ^2=8.50 with p<0.005 for upvotes and χ^2=6.77 with p<0.01 for downvotes).

UGL participants were better able to find diverse rationale for others' upvotes and downvotes (M:5.82, SD:1.05 for upvotes and M:5.62, SD:1.22 for downvotes) than Binary participants (M:4.99, SD:1.61 for upvotes and M:4.72, SD:1.67 for downvotes). The differences were significant (MW test, Z=4.08 for upvotes and Z=4.08 for downvotes, both with p<0.0001). UGL participants liked how diverse reasons for up/down voting a comment are expressed through UGLs. One UGL participants said, *"It's nice to see why another person reacted to a comment. For instance, was it because the commenter was rude or was it that they were misleading or was there just a disagreement on the entire subject premise."*

5.3.4 ***RQ4**: How do UGLs affect participants' tolerance to the opinions that do not align with theirs?* In both pre- and post-surveys, participants expressed moderate tolerance towards people with opposite stance on each topic. For both UGL (M: 4.54, SD: 1.44 for pre-survey and M:4.53, SD: 1.48 for post-survey) and Binary (M: 4.76, SD: 1.30 for pre-survey and M:4.74, SD: 1.45 for post-survey) conditions, participants' level of tolerance did not significantly change after the main activity. Likewise, participants in both conditions showed moderate tolerance towards others' reactions that do not align with theirs in both pre-and post-survey. For both UGL (M: 4.70, SD: 1.22 and M:4.64, SD: 1.21 for pre- and post-survey) and Binary (M: 4.84, SD: 1.15 and M:4.77, SD: 1.22 for pre- and post-survey) conditions, there was no significant difference between participants' tolerance level before and after the main activity.

Although there was no significant change in the tolerance towards people with opposite stances on the topic, UGL participants left more positive reactions to comments with opposite stances from theirs. Among three initial comments with opposite stances from the participant, UGL participants left positive reactions to 1.23 (SD: 1.55) comments on average. On the other hand, Binary participants left positive reactions to 0.58 (SD: 0.87) out of three comments that had opposite stances on the topic. The difference was statistically significant (MW test, Z=2.74 with p<0.05). One participant in UGL condition also noted that *"I like that there is reasoning behind the upvoting and downvoting process rather than just a simple 'I like it or I don't'. There are things that you can disagree with but are very well written and logical, and it's great to be able to convey that. ··· "*

6 DISCUSSION

Our evaluation showed that UGLs help users better express their thoughts and understand the multifacetedness of people's reaction to a comment. In this section, we revisit our findings and refine the design implications of our current study on 1) the information rich-reach trade-off of reactions in the comment section and 2) the effect of capturing diverse reactions through UGLs on tolerance towards the outgroup.

6.1 Information Rich-Reach Trade-off.

While generating new UGLs requires more mental and physical effort, UGLs mitigate the information-rich-reach trade-off in a comment section. UGLs are information-rich, reflecting nuanced differences across people's viewpoints about a comment. Replacing up/downvotes with UGLs did not reduce the overall level of clicks. In fact, the number of votes with UGLs eventually surpassed the number of up/downvotes across all four topic conditions.

Our findings imply that facilitating open evaluation is key to active participation in the UGL condition. UGLs enabled participants with moderate or mixed opinions to express their opinions without projecting them onto up/downvotes. Moreover, participants leaving reactions to comments felt that they were able to contribute more through UGLs. This aligns with previous studies that report how the perceived uniqueness of individual contributions may increase user participation [4, 15, 22].

As we observed in the study, however, the success of UGLs depends on how early users engage with the system. The benefits of having UGLs become pronounced when there exists a certain number of UGLs that users can engage with. This observation is in line with previous work on how initial contributions (e.g., first comment [47] or first review [7]) govern the success of an online community. Systemic support that can reduce early users' burden of generating UGLs, e.g., suggesting a list of potential UGLs that users can refer to, can be introduced to accelerate the generation of UGLs.

6.2 The Effect of Showing Diverse Reactions through UGLs.

We found that presenting the rich context of user reactions to comments affects how readers understand and evaluate each comment. In response to the open-ended question about their experience with UGLs, participants noted that UGLs helped them evaluate a comment more thoroughly. For example, one user mentioned that she re-considered a comment that she had originally thought was good because she found a UGL pointing its inaccuracy. We found that this mechanism had a positive effect in encouraging more deliberate participation across users.

The effect of diversity on reducing affective polarization has gained scholarly attention [1, 3, 17, 23, 30, 37, 42]. While aggregated UGLs improved participants' understanding of others' multifaceted evaluation of a comment, our findings show that simply increasing exposure to diversity does not increase the tolerance towards the outgroup. Participants in our study noted that seeing other's UGL that aligns with their opinion increased their confidence. It could be that, others' endorsement of labels they agree with 1) reinforced their attitudes and beliefs and 2) offset the positive effect of exposure to diverse opinions from the other side. Recent research reports similar findings from exploring the effect of seeing diverse opinions on social issues [17]. Nevertheless, participants' qualitative responses suggest that UGLs could still provide a starting point for a more productive conversation in the future. One participant pointed out that UGLs helped them understand how others' evaluations of a comment are different from their own. This implies that UGLs can mitigate misunderstanding and misinterpretation over aggregated up/downvotes.

7 LIMITATIONS AND FUTURE WORK

Our current study has several limitations, which leave room for future work. In our study, we only look at interactions between readers and reactors. The complete mechanism of the comment section involves commenters, readers, and reactors, though one user can take multiple roles. Future studies could further explore whether UGLs have positive effects on subsequent comment writing behaviors of reactors and parent commenters. Increased richness in feedback from users could motivate commenters to be more deliberate and leave higher quality comments, which can lead to higher quality replies [46].

Furthermore, future work should conduct a longitudinal deployment study to re-examine the effect of UGLs. Our study design and setting (having crowd workers as participants or assigning a topic they discuss) limit the generalizability of our study findings. Also, there could have been a novelty effect of UGLs, likely leading to more active usage. A longer-term observation of how UGLs would work in real-world settings could help us validate our results.

Last but not least, we could explore the application of evaluative UGLs in a different setting other than commenting context on serious or controversial current events. For example, one could investigate how UGLs could improve user experience on community platforms focused on asking and answering questions.

8 CONCLUSION

In this work, we identified limitations of up/downvotes and proposed user-generated labels (UGLs) as an alternative reaction design that captures diverse and precise reactions to a comment. UGLs leverage social-psychological incentives and enables the production of a more nuanced, rich picture of a comment's value to other users. Our evaluation results demonstrated that users were more expressive and left more reactions to comments with UGLs than with up/downvotes. We also show that UGLs help users interpret others' reactions and understand the multifacetedness of people's reactions to comments. We anticipate that our design of UGL and study findings can guide and inspire the future design of reactions to better capture and deliver users' thoughts on comments.

9 ACKNOWLEDGEMENTS

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00075, Artificial Intelligence Graduate School Program (KAIST)) and KAIST new faculty research fund (G04200008). Jeong-woo Jang and Juho Kim are co-corresponding authors.

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 (July 2018), 964–981. https://doi.org/10.1086/697253

[2] Tanja Aitamurto. 2015. Motivation Factors in Crowdsourced Journalism: Social Impact, Social Change, and Peer Learning. *International Journal of Communication* 9 (Oct. 2015). https://ijoc.org/index.php/ijoc/article/view/3481

[3] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (Sept. 2018), 9216–9221. https://doi.org/10.1073/pnas.1804840115

[4] Gerard Beenen, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut. 2004. Using Social Psychology to Motivate Contributions to Online Communities. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. ACM, New York, NY, USA, 212–221. https://doi.org/10.1145/1031607.1031642

[5] Kimmy Wa Chan and Stella Yiyan Li. 2010. Understanding consumer-to-consumer interactions in virtual communities: The salience of reciprocity. *Journal of Business Research* 63, 9-10 (Sept. 2010), 1033–1040. https://doi.org/10.1016/j.jbusres.2008.08.009

[6] Coye Cheshire. 2007. Selective Incentives and Generalized Information Exchange. *Social Psychology Quarterly* 70, 1 (March 2007), 82–100. https://doi.org/10.1177/019027250707000109

[7] Dan Cosley, Dan Frankowski, Sara Kiesler, Loren Terveen, and John Riedl. 2005. How Oversight Improves Member-Maintained Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 11–20. https://doi.org/10.1145/1054972.1054975

[8] W Phillips Davison. 1983. The Third-Person Effect in Communication. *Public opinion quarterly* 47, 1 (Jan. 1983), 1–15. https://doi.org/10.1086/268763

[9] Chrysanthos Dellarocas, Guodong Gao, and Ritu Narayan. 2010. Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products? *Journal of Management Information Systems* 27, 2 (2010), 127–158. https://doi.org/10.2753/MIS0742-1222270204

[10] Charles A. Dorison, Julia A. Minson, and Todd Rogers. 2019. Selective exposure partly relies on faulty affective forecasts. *Cognition* 188 (July 2019), 98–107. https://doi.org/10.1016/j.cognition.2019.02.010

[11] James N. Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2021. Affective polarization, local contexts and public opinion in America. *Nature Human Behaviour* 5, 1 (Jan. 2021), 28–38. https://doi.org/10.1038/s41562-020-01012-5

[12] James N Druckman and Matthew S Levendusky. 2019. What Do We Measure When We Measure Affective Polarization? *Public Opinion Quarterly* 83, 1 (May 2019), 114–122. https://doi.org/10.1093/poq/nfz003

[13] Heinz Holländer. 1990. A Social Exchange Approach to Voluntary Cooperation. *American Economic Review* 80, 5 (Dec. 1990), 1157–1167. https://www.jstor.org/stable/2006767

[14] Shih-Wen Huang and Wai-Tat Fu. 2013. Don't Hide in the Crowd! Increasing Social Transparency between Peer Workers Improves Crowdsourcing Outcomes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 621–630. https://doi.org/10.1145/2470654.2470743

[15] Steven J Karau and Kipling D Williams. 1993. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology* 65, 4 (Oct. 1993), 681–706. https://doi.org/10.1037/0022-3514.65.4.681

[16] Eunkyung Kim, Dietram Scheufele, and Jeong Yeob Han. 2011. Structure or Predisposition? Exploring the Interaction Effect of Discussion Orientation and Discussion Heterogeneity on Political Participation. *Mass Communication and Society* 14, 4 (2011), 502–526. https://doi.org/10.1080/15205436.2010.513469

[17] Hyunwoo Kim, Haesoo Kim, Kyung Je Jo, and Juho Kim. 2021. StarryThoughts: Facilitating Diverse Opinion Exploration on Social Issues. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 66 (April 2021), 29 pages. https://doi.org/10.1145/3449140

[18] Yonghwan Kim. 2015. Does Disagreement Mitigate Polarization? How Selective Exposure and Disagreement Affect Political Polarization. *Journalism & Mass Communication Quarterly* 92, 4 (Dec. 2015), 915–937. https://doi.org/10.1177/1077699015596328

[19] Ezra Klein. 2020. *Why we're polarized.* Avid Reader Press, New York, NY, USA.

[20] Eun-Ju Lee and Yoon Jae Jang. 2010. What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research* 37, 6 (Dec. 2010), 825–846. https://doi.org/10.1177/0093650210376189

[21] Shu-Yueh Lee, Sara Steffes Hansen, and Jin Kyun Lee. 2016. What makes us click "like" on Facebook? Examining psychological, technological, and motivational factors on virtual endorsement. *Computer Communications* 73 (Jan. 2016), 332–341. https://doi.org/10.1016/j.comcom.2015.08.002

[22] Pamela J. Ludford, Dan Cosley, Dan Frankowski, and Loren Terveen. 2004. Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. ACM, New York, NY, USA, 631–638. https://doi.org/10.1145/985692.985772

[23] Nahema Marchal. 2021. "Be Nice or Leave Me Alone": An Intergroup Perspective on Affective Polarization in Online Political Discussions. *Communication Research* (Sept. 2021). https://doi.org/10.1177/00936502211042516

[24] Gina M. Masullo and Jiwon Kim. 2021. Exploring "Angry" and "Like" Reactions on Uncivil Facebook Comments That Correct Misinformation in the News. *Digital Journalism* 9, 8 (Sept. 2021), 1103–1122. https://doi.org/10.1080/21670811.2020.1835512

[25] Douglas M McLeod, Benjamin H Detenber, and William P Eveland Jr. 2001. Behind the third-person effect: Differentiating perceptual processes for self and other. *Journal of Communication* 51, 4 (Dec. 2001), 678–695. https://doi.org/10.1111/j.1460-2466.2001.tb02902.x

[26] Y Yi Mun and Yujong Hwang. 2003. Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies* 59, 4 (Oct. 2003), 431–449. https://doi.org/10.1016/S1071-5819(03)00114-9

[27] German Neubaum and Nicole C. Krämer. 2017. Monitoring the Opinion of the Crowd: Psychological Mechanisms Underlying Public Opinion Perceptions on Social Media. *Media Psychology* 20, 3 (July 2017), 502–531. https://doi.org/10.1080/15213269.2016.1211539

[28] Elisabeth Noelle-Neumann. 1974. The Spiral of Silence A Theory of Public Opinion. *Journal of Communication* 24, 2 (June 1974), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

[29] Blair Nonnecke and Jenny Preece. 2000. Lurker Demographics: Counting the Silent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. ACM, New York, NY, USA, 73–80. https://doi.org/10.1145/332040.332409

[30] Thomas F. Pettigrew and Linda R. Tropp. 2006. A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology* 90, 5 (2006), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

[31] Pablo Porten-Cheé and Christiane Eilders. 2020. The effects of likes on public opinion perception and personal opinion. *Communications* 45, 2 (2020), 223–239. https://doi.org/10.1515/commun-2019-2030

[32] Jenny Preece, Blair Nonnecke, and Dorine Andrews. 2004. The top five reasons for lurking: improving community experiences for everyone. *Computers in human behavior* 20, 2 (March 2004), 201–223. https://doi.org/10.1016/j.chb.2003.10.015

[33] Emily Pronin, Thomas Gilovich, and Lee Ross. 2004. Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. *Psychological Review* 111, 3 (2004), 781–799. https://doi.org/10.1037/0033-295X.111.3.781

[34] Tripti Ghosh Sharma, Juho Hamari, Ankit Kesharwani, and Preeti Tak. 2020. Understanding continuance intention to play online games: roles of self-expressiveness, self-congruity, self-efficacy, and perceived risk. *Behaviour & Information Technology* (2020), 1–17. https://doi.org/10.1080/0144929X.2020.1811770

[35] David K. Sherman, Leif D. Nelson, and Lee D. Ross. 2003. Naïve Realism and Affirmative Action: Adversaries are More Similar Than They Think. *Basic and Applied Social Psychology* 25, 4 (Dec. 2003), 275–289. https://doi.org/10.1207/S15324834BASP2504_2

[36] Jane B Singer. 2014. User-generated visibility: Secondary gatekeeping in a shared media space. *New Media & Society* 16, 1 (Feb. 2014), 55–73. https://doi.org/10.1177/1461444813477833

[37] Jonathan Stray. 2021. Designing Recommender Systems to Depolarize. https://doi.org/10.48550/arXiv.2107.04953 arXiv:2107.04953 [cs.IR]

[38] Natalie Jomini Stroud, Ashley Muddiman, and Joshua M Scacco. 2017. Like, recommend, or respect? Altering political behavior in news comment sections. *New Media & Society* 19, 11 (Nov. 2017), 1727–1743. https://doi.org/10.1177/1461444816642420

[39] Natalie Jomini Stroud, Emily Van Duyn, and Cynthia Peacock. 2016. News commenters and news comment readers. *Engaging News Project* (2016), 1–21.

[40] Erin M Sumner, Luisa Ruge-Jones, and Davis Alcorn. 2018. A functional approach to the Facebook Like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. *New Media & Society* 20, 4 (April 2018), 1451–1469. https://doi.org/10.1177/1461444817697917

[41] S S Sundar and C Nass. 2001. Conceptualizing Sources in Online News. *Journal of Communication* 51, 1 (March 2001), 52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

[42] Charles S. Taber and Milton Lodge. 2006. Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science* 50, 3 (July 2006), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

[43] Ye Tian, Thiago Galery, Giulio Dulcinati, Emilia Molimpakis, and Chao Sun. 2017. Facebook sentiment: Reactions and Emojis. In *Proceedings of*

the Fifth International Workshop on Natural Language Processing for Social Media. Association for Computational Linguistics, Valencia, Spain, 11–16. https://doi.org/10.18653/v1/W17-1102

[44] Joseph B. Walther and Jeong-woo Jang. 2012. Communication Processes in Participatory Websites. *Journal of Computer-Mediated Communication* 18, 1 (Oct. 2012), 2–15. https://doi.org/10.1111/j.1083-6101.2012.01592.x

[45] Joseph B. Walther, Jeong-woo Jang, and Ashley A. Hanna Edwards. 2018. Evaluating Health Advice in a Web 2.0 Environment: The Impact of Multiple User-Generated Factors on HIV Advice Perceptions. *Health Communication* 33, 1 (Jan. 2018), 57–67. https://doi.org/10.1080/10410236.2016.1242036

[46] Yixue Wang and Nicholas Diakopoulos. 2021. The role of New York times picks in comment quality and engagement. In *Proceedings of the 54th Annual Hawaii International Conference on System Sciences (HICSS 2021)*. IEEE Computer Society, 2924–2933.

[47] Tim Weninger, Xihao Avi Zhu, and Jiawei Han. 2013. An exploration of discussion threads in social news sites: A case study of the Reddit community. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. 579–583. https://doi.org/10.1145/2492517.2492646

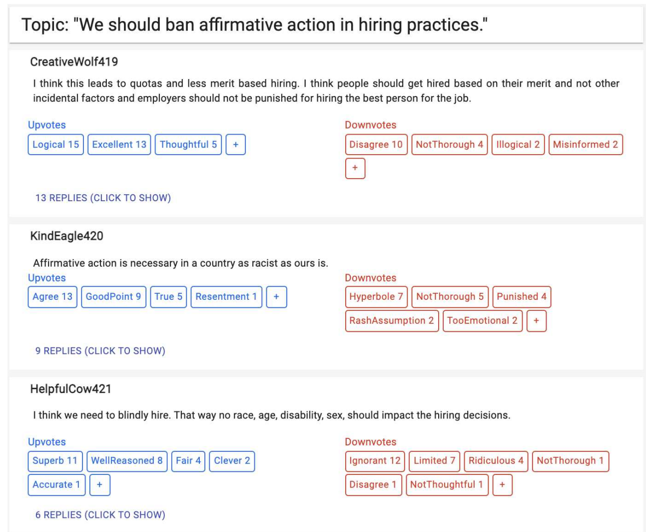
[48] Thomas Zerback, Thomas Koch, and Benjamin Krämer. 2015. Thinking of Others: Effects of Implicit and Explicit Media Cues on Climate of Opinion Perceptions. *Journalism & Mass Communication Quarterly* 92, 2 (June 2015), 421–443. https://doi.org/10.1177/1077699015574481

A NUMBER OF GENERATED COMMENTS, REPLIES, UGLS, AND VOTES

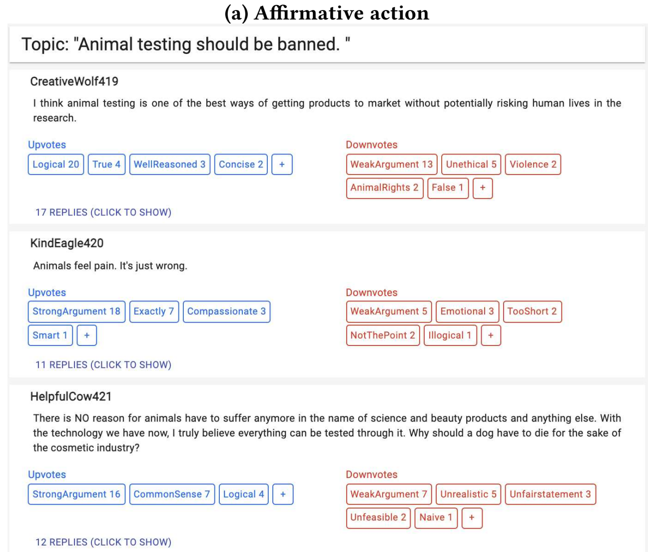
Table 2: Number of participants, comments, replies, UGLs, and votes generated by topic and condition.

Topic	Condition	# participants	# comments	# replies	# UGLs	# unique UGLs	# votes
Capital punishment	Binary	28	16	65	-	-	247
	UGL	29	17	58	85	49	369
Affirmative action	Binary	24	13	53	-	-	241
	UGL	25	10	40	69	50	299
Animal testing	Binary	26	16	47	-	-	249
	UGL	28	15	82	142	92	488
Consumer data	Binary	31	9	59	-	-	290
	UGL	27	14	85	98	72	474
Total	Binary	109	54	224	-	-	1027
	UGL	109	56	265	394	234	1630

B DISCUSSION INTERFACE WITH UGLS



(a) Affirmative action



(b) Animal testing

Figure 5: Screenshots of the interface with UGLs generated in the discussion on (a) affirmative action and (b) animal testing topics.