Intergroup Contact in the Wild: Characterizing Language Differences between Intergroup and Single-group Members in NBA-related Discussion Forums

JASON SHUO ZHANG, University of Colorado Boulder, USA
CHENHAO TAN, University of Colorado Boulder, USA
QIN LV, University of Colorado Boulder, USA

Intergroup contact has long been considered as an effective strategy to reduce prejudice between groups. However, recent studies suggest that exposure to opposing groups in online platforms can exacerbate polarization. To further understand the behavior of individuals who actively engage in intergroup contact in practice, we provide a large-scale observational study of intragroup behavioral differences between members with and without intergroup contact. We leverage the existing structure of NBA-related discussion forums on Reddit to study the context of professional sports. We identify fans of each NBA team as members of a group and trace whether they have intergroup contact. Our results show that members with intergroup contact use more negative and abusive language in their affiliated group than those without such contact, after controlling for activity levels. We further quantify different levels of intergroup contact and show that there may exist nonlinear mechanisms regarding how intergroup contact relates to intragroup behavior. Our findings provide complementary evidence to experimental studies in a novel context, and also shed light on possible reasons for the different outcomes in prior studies.

CCS Concepts: • **Applied computing → Law, social and behavioral sciences**; • **Human-centered computing → Collaborative and social computing**.

Additional Key Words and Phrases: intergroup contact, polarization, intragroup behavior, language usage, NBA-related discussion forums

ACM Reference Format:
Jason Shuo Zhang, Chenhao Tan, and Qin Lv. 2019. Intergroup Contact in the Wild: Characterizing Language Differences between Intergroup and Single-group Members in NBA-related Discussion Forums. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW, Article 193 (November 2019), 35 pages. https://doi.org/10.1145/3359295

1 INTRODUCTION

Driven by the growing concerns of tribalism and polarization in world politics [16, 45, 96], it is increasingly important to understand intergroup contact as a straightforward yet potentially powerful strategy to reduce prejudice between groups. Intergroup contact refers to interactions

We thank anonymous reviewers and members of the NLP+CSS research group at CU Boulder for their insightful comments and discussions; Maria Deslis for illustrating Figure 1; Scott Fredrick Holman from the CU Boulder Writing Center for his feedback and support during the writing process; Jason Baumgartner for sharing the dataset that enabled this research. This work is supported in part by the US National Science Foundation (NSF) through grant CNS 1528138.
Authors' addresses: Jason Shuo Zhang, jasonzhang@colorado.edu, University of Colorado Boulder, Boulder, CO, 80309, USA; Chenhao Tan, chenhao@chenhaot.com, University of Colorado Boulder, Boulder, CO, 80309, USA; Qin Lv, qin.lv@colorado.edu, University of Colorado Boulder, Boulder, CO, 80309, USA.

2573-0142/2019/11-ART193 $15.00
https://doi.org/10.1145/3359295

between members of different groups, and groups can be defined using a variety of factors, including political ideology, place of origin, and race. A key hypothesis is that members with intergroup contact (henceforth "intergroup members") behave differently, e.g., by showing sympathy towards other groups and voicing different opinions in their affiliated group [2, 3, 28, 29, 31, 52, 71, 78–81]. However, prior studies have observed different effects of intergroup contact. For instance, self-reported surveys show that intergroup contact relates to reduced prejudice towards immigrants in European countries [78], while a recent experimental study finds that exposure to opposing groups on Twitter can exacerbate political polarization [3]. Although self-reported surveys and experimental studies have been the main methods for studying the effect of intergroup contact [2, 8, 10, 34, 52, 71, 73], we believe that observational study allows researchers to characterize intergroup contact in the wild and provide valuable complementary evidence in diverse contexts. Indeed, with the emergence of online groups, it has become possible to observe intergroup contact and individual behavior at a massive scale for substantial periods. Our goal in this study is to investigate how individuals that choose to engage in intergroup contact behave differently from others without intergroup contact in their original affiliated group in online platforms.

We leverage the existing structure of NBA-related discussion forums on Reddit to identify the group affiliations of users and intergroup contact in the context of professional sports, a novel domain different from politics. We choose online fan groups of professional sports teams as a testbed for the following reasons. First, professional sports play a significant role in modern life [11, 40, 106]. People in the United States spent more than 31 billion hours watching sports games in 2015 [70], and the attendance of the 2017-2018 National Basketball Association (NBA) season reached 22 million [69]. Second, professional sports teams are unambiguously competitive in nature. Similar to other common contexts for studies on intergroup contact (e.g., political ideology), fans of sports teams can treat fans of opposing teams as enemies and sometimes even engage in violence [30, 85]. Moreover, sports fans tend to think that the media and supporters from opposing teams are likely to have unfair opinions against their favored teams, just like people with different ideologies [26].

Figure 1 illustrates our framework. There are 30 teams in the NBA, and every team has its discussion forum (henceforth *team subreddit*) on Reddit, a place where fans of the corresponding team congregate and discuss news, games, and any other topics that are relevant to the team. The low-access barrier on the Internet also enables users to communicate easily with fans from opposing teams. In fact, /r/NBA is dedicated to interactions between fans of all NBA teams for any discussion related to the NBA. Contrary to each team's "echo chamber," which is dominated by fans that support the same team, /r/NBA represents an open and diverse environment where intergroup contact occurs. We can thus identify intergroup members and single-group members based on whether they have any activity in the intergroup setting (/r/NBA).

As posting comments is a major activity in online platforms such as Reddit, analyzing the language used in online platforms provides an opportunity for capturing individuals' attitudes and emotions [24, 75, 82, 88]. In particular, there has been growing concerns about hate speech and negative language in online communities [12, 13, 15, 23, 89]. We thus focus on characterizing the differences in language usage between intergroup and single-group members in *their affiliated team subreddit* (the intragroup setting), e.g., whether intergroup members swear more than single-group members in their affiliated team subreddit. As a result, we would be able to capture behavioral differences reflected in language use between intergroup and single-group members in the intragroup setting. Note that in this work, we do not claim that intergroup contact causes such differences due to endogenous factors that may lead to individuals choosing to engage in intergroup contact in the wild (i.e., individuals who choose to engage in intergroup contact in practice may be inherently different from those who do not).

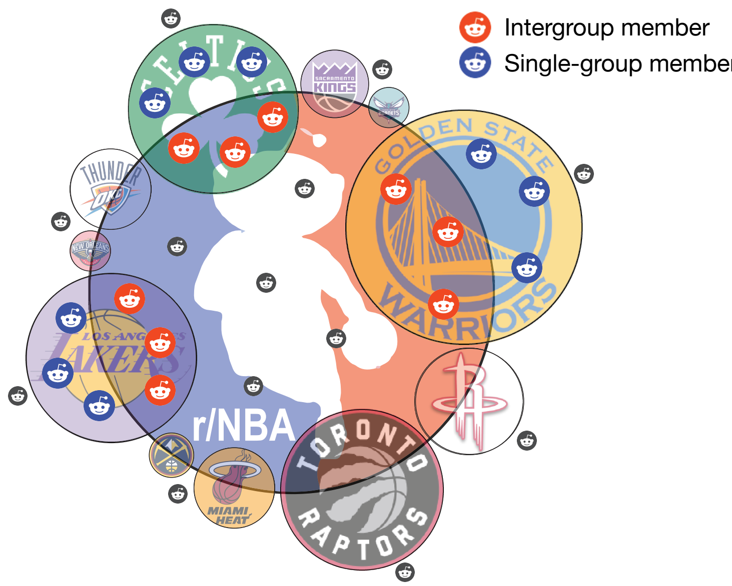


Fig. 1. Illustration of NBA-related discussion forums (also known as subreddits) on Reddit. We identify group affiliation (i.e., whether a person is a fan of an NBA team) and intergroup contact based on the existing structure of NBA-related subreddits on Reddit. Each team has its team subreddit. Here, we present 11 of the 30 NBA teams (with corresponding team logos) to cover subreddits of different sizes. The central /r/NBA logo represents /r/NBA, where intergroup contact happens. The radius of each logo is proportional to the number of subscribers in the corresponding subreddit. Users in each team logo represent fans of a team based on their indicated support in NBA-related subreddits. Red icons refer to intergroup members who have engaged in intergroup contact and are thus also in the /r/NBA logo, while blue icons refer to single-group users without such behavior. We only put red and blue icons in the three largest team subreddits due to space limitations, but every team subreddit has these two categories of users (see Figure 4 for the number of intergroup and single-group members of each team). Note that not all users who participated in these discussion forums qualify as a fan of an NBA team (grey icons).

Organization and highlights. We start by summarizing related work to put our work in context (Section 2). We then introduce our dataset and provide an overview of the framework for identifying group affiliations and intergroup contact in Section 3. With intergroup members and single-group members identified, We investigate two research questions in the rest of the paper (methods in Section 4 and results in Section 5):

RQ1: *How do members with intergroup contact differ from those without such contact in intragroup language usage in NBA fan groups?*

RQ2: *How do different levels of intergroup contact relate to intragroup language usage?*

For **RQ1**, we first apply matching techniques to make sure the intergroup and single-group members are comparable. We then analyze the behavioral differences between intergroup and single-group members by examining language usage of their comments in their affiliated team subreddit. We demonstrate intriguing contrasts between them: intergroup members tend to use more negative and swear words, and generate more hate speech comments compared to single-group members in their affiliated team subreddit.

For **RQ2**, we are able to quantify different levels of intergroup contact for each intergroup member based on the frequency of intergroup contact. Interestingly, we find varying mechanisms of how different levels of intergroup contact relate to intragroup behavioral differences. For instance, although intergroup contact mostly monotonically relates to differences in language usage, the trends are not necessarily linear. Such varying mechanisms provide complementary evidence to the seemingly conflicting results on intergroup contact in recent studies.

To explore the potential reasons behind the clear behavioral differences in language usage between intergroup and single-group members, we further compare the language usage of intergroup members between the intragroup setting (affiliated team subreddit) and the intergroup setting (/r/NBA) in Section 6. This setup naturally controls for the subject because we compare the same person across two different environments. We find that intergroup members are even more negative and more likely to swear in the intergroup setting. Such negative intergroup contact may partly explain the observed differences in intragroup language usage.

Our work highlights the fact that individuals selectively choose to have intergroup contact in the wild, and in turn interact with people without intergroup contact in their original group. We further demonstrate a variety of ways in which intergroup contact levels can moderate intragroup behavior. These observations may reconcile recent conflicting results with respect to intergroup contact. Our findings indicate that observational studies can provide important complementary evidence to experimental studies on this topic because interventions can hardly result in deep and regular contacts. We offer discussions in Section 7 and conclude our work in Section 8.

2 RELATED WORK

In this section, we first discuss studies that use language as a lens to understand human behavior, especially recent studies on the use of negative language in the context of antisocial behavior. Next, we explain the growing concerns of tribalism, echo-chambers, and polarization, and highlight our specific context, sports, as a testbed for understanding these issues. We then discuss the role of intergroup contact in affecting individual opinions towards opposing groups, including recent work on its backfire effect in online platforms. Finally, we point out opportunities in online sports discussion forums for understanding human behavior.

2.1 Language as a Lens of Human Behavior

The proliferation of textual content online has inspired a vast body of literature to understand the language in online communication and its relationship with individual attributes. Prior research in CSCW and related communities has investigated how language can reflect properties of individuals [24, 68, 75, 82, 88, 100, 102]. For instance, Toma and Hancock [102] show that linguistic emotions correlate with deception in online dating profiles; De Choudhury et al. [24] uses linguistics style features to show that mothers with post-partum depression are more likely to use first-person singular pronouns and swear words; Naaman et al. [68] conduct a quantitative analysis of message content from over 350 Twitter users to characterize the type of messages posted on the platform and broadly classify users as self-broadcasters and informers. In general, users' demographic information and personality can also be predicted based on linguistic features extracted from textual social media data [32, 59, 63, 64].

Recently, negative language use has been examined in the context of antisocial behavior in online communities [7, 12–15, 23, 89]. Conceptually related is a prior study on the effects of community feedback on user behavior, which reveals that negative feedback can lead to future antisocial behavior [14]. Cheng et al. [13] further design an experiment that shows negative mood expressed from textual content increases the likelihood of trolling in online platforms. A supervised learning model proposed by Blackburn and Kwak [7] indicates that negative sentiments are useful for predicting toxic players in online games. In this work, we compare the differences in language usage between intergroup and single-group members, with a focus on expressions of emotions.

2.2 Tribalism, Echo Chambers, and Polarization

A battery of studies in social sciences has shown that human behavior is shaped by our need to belong to a group and by our proclivity to hate rival groups [17, 79, 98]. Such behavior has been documented in a wide variety of contexts. In the political context, for instance, recent studies find that "liberal group" and "conservative group" on social media not only rarely talk to each other, but also use different hashtags and links to various websites within their tweets [35, 37, 60, 76, 92]. Another commonly studied context is brand communities in the marketing literature [6, 20, 38, 44, 67]. For example, Hickman and Ward [44] show that in-groups are strongly motivated to develop negative views of out-groups and engage in "trash talking" about out-groups. In-group members will also gain pleasure at the misfortune of rival brands and their users.

In the sports context, the team sports literature focus on the negative consequences of rivalry, such as negative explicit and implicit attitudes towards the opposing team [17, 53], *schadenfreude* [18, 41], and even riot [39]. These negative perceptions may even transfer to the sponsors of the rival team: Dalakas and Levin [21] explore the negative sponsorship effects and find that sponsors of disliked NASCAR drivers are viewed less positively than sponsors of liked drivers. Similarly, Olson [72] finds that brands faced a steep decline in sales among Manchester City fans when they announced the sponsorship of the soccer club Manchester United, a fierce rival of Manchester City. By examining the intergroup emotions of fans of the Boston Red Sox and New York Yankees, Lehr et al. [53] show that pleasure from a powerful rival's losses can outstrip that from gains of the supported team. Given the competitive nature of professional sports and the importance of emotions in fan behavior, we believe that professional sports provide exciting opportunities for understanding polarization.

2.3 Intergroup Contact

Intergroup contact has long been considered as an effective strategy to reduce prejudice between groups [29]. For instance, a seminal work by Pettigrew [78] shows that intergroup contact relates to reduced prejudice towards immigrants based on self-reported surveys in France, Great Britain, the Netherlands, and West Germany. Wright et al. [109] find correlational evidence that people who knew that an in-group member had an out-group friend had less negative intergroup attitudes. They also experimentally demonstrate that providing this information induces more positive attitudes. Abbott and Cameron [1] examine young people's assertive bystander intentions in an intergroup (immigrant) name-calling situation and find that greater intergroup contact is related to higher levels of empathy, higher levels of cultural openness, and reduced intergroup bias. From the perspective of language usage online, a field experiment designed by White et al. [107] demonstrates that Muslim and Christian high-school students who have structured Internet intergroup interactions tend to use more affective and positive emotion words, and less anger and sadness words. Kim and Wojcieszak [50] test online contact with two distinct out-groups, undocumented immigrants and gay people. They find that direct online contact improves attitudes towards both out-groups through positive and negative emotions, whereas extended online contact reduces negative emotions and improve attitudes towards undocumented immigrants.

However, recent studies on the "backfire" effect suggest that exposure to opposing groups in online platforms can exacerbate political polarization [2, 3, 52, 71]. [1] For instance, Bail et al. [3] introduce intergroup contact by following a Twitter bot that aggregates tweets of opinion leaders

[1] Wood and Porter [108] show that backfire in Nyhan and Reifler [71] is stubbornly difficult to reproduce, which further demonstrates the varying results in recent studies.

from the opposing political ideology and find that Republicans who follow a liberal Twitter bot become substantially more conservative. Lee et al. [52] use panel data collected in South Korea to investigate the effects of social media usage on changing the political view. They highlight the role of social media in activating political participation and pushing users toward ideological poles.

A possible way to reconcile such differences in prior literature is to review the mechanisms that contribute to the positive effects of intergroup contact: (1) enhancing knowledge about other groups, (2) reducing anxiety when facing opposing groups, and (3) increasing empathy and perspective-taking [28, 29, 78–81, 109]. Depending on the motivations to engage in intergroup contact and the actual activities during the contact, intergroup contact in online platforms may not necessarily achieve these goals. We aim to conduct a large-scale observational study to understand the differences between intergroup and single-group members in their original affiliated group, and also provide some insights on the nature of intergroup contact in /r/NBA.

It is useful to point out that there is little work on intergroup contact in the CSCW community. In the meanwhile, several recent studies provide a characterization of intergroup conflict. Kumar et al. [51] examine cases of intergroup conflict across 36,000 communities on Reddit where users of one community are mobilized by negative sentiment to comment in another community and show that less than 1% communities start 74% conflicts. At the community level, by constructing a conflict network between subreddits, Datta and Adar [22] find that larger subreddits are more likely to be involved in conflicts with a large number of subreddits, and the main "targets" change over time. However, intergroup contact is different from intergroup conflict as it may help improve intergroup attitudes and reduce intergroup tensions and conflicts. Approaching this topic from a CSCW lens raises additional questions about how socio-technical design decisions can influence the outcomes reported in traditional offline settings.

2.4 Fan Behavior in Online Sports Communities

Despite the important role of sports in modern life, sports fan behavior in online communities for professional sports remains understudied. Yu and Wang [112] collect real-time tweets from US soccer fans during five 2014 FIFA World Cup games to examine soccer fans' emotional responses in their tweets. The quantitative analyses show that fear and anger were the most common negative emotions and in general increased when the opponent team scored and decreased when the US team scored. Zhang et al. [113] investigate the connection between online fan behavior and offline team performance, and Leung et al. [54] study the effect of NFL game outcomes on content contribution to Wikipedia. It is also worth noting that fan behavior can differ depending on the environment. Cottingham [19] demonstrates the difference in emotional energy between fans in sports bars and those attending the game in the stadium. We believe that there exist exciting opportunities in online sports discussion forums for understanding human behavior, including intergroup contact.

3 PROFESSIONAL SPORTS AS A TESTBED

We focus on the professional sports context derived from NBA-related discussion forums (/r/NBA and 30 team subreddits) on Reddit, an active community-driven platform where users can submit posts and make comments. These user-created discussion forums are also called "subreddits". Each subreddit has multiple moderators to make sure that posts are relevant to the subreddit's theme. Over the years, basketball fans all over the world have flocked to /r/NBA, the site's professional basketball subreddit, to discuss games in progress, seek meaning in the latest trade rumors, and debate the legality of calls by the referees. In fact, /r/NBA has become the largest single-sport subreddit with more than 1.9M subscribers [84] and one of the most active subreddits on Reddit [83]. NBA-related subreddits represent an ideal testbed for understanding how intergroup contact relates to intragroup behavior because their structure allows us to identify many users' team affiliation.

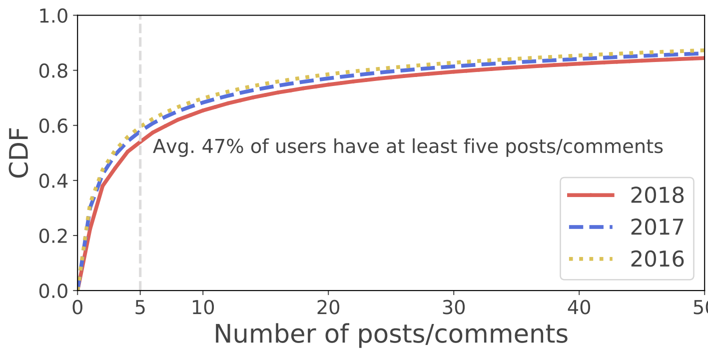


Fig. 2. The distribution of the number of activities made by users in NBA-related subreddits in the 2018, 2017, and 2016 seasons.

Moreover, /r/NBA is the place for all basketball fans to congregate, where intergroup contact between fans of different teams happens.

3.1 Dataset and NBA Seasons

We obtain 2.1M posts and 61M comments in NBA-related subreddits from https://pushshift.io [5]. As pointed out in Zhang et al. [113], offline NBA seasons are reflected in user behavior in these NBA-related subreddits. We organize our dataset according to the timeline of NBA seasons and focus on the most recent three seasons, i.e., from July 2015 to June 2018. For simplicity and clarity, we refer to a specific season by the calendar year when it ends. For instance, the official 2017-2018 NBA season is referred to as *the 2018 season* or *2018* in this paper.

3.2 Identifying Team Affiliation and Intergroup Contact

To identify the team affiliation of users in a season, we first define active users in NBA-related subreddits in a season as those who have at least five activities, where an activity refers to either submitting a post or making a comment. Figure 2 shows the distribution of the number of activities by a user in NBA-related subreddits. These active users contribute over 95% of all the activities in NBA-related subreddits.

We identify the team affiliation of active users based on where their activities occur and by using a special mechanism on Reddit, known as flair. Flair appears as an icon next to the username in posts and comments. Every comment can have at most one flair. Before April 2018, flairs are represented by team logos. After that, Reddit adopted a new design to the entire platform, and the flairs are represented by team names in /r/NBA. An example is shown in Figure 3. In /r/NBA, fans can use flairs to indicate support of a team. ~80% of the comments/posts in our /r/NBA dataset have been made with flairs even though flairs are optional. We use all the flairs that fans used in /r/NBA for the inference of their team affiliation [2].

We view posting/commenting in a team subreddit and using a team's flair in /r/NBA as an indication of support towards that team. An active user is defined as a fan of a team if the user indicates support only for that team and such support sustains over all activities in an entire NBA

[2]In our /r/NBA dataset, every comment's JSON format has the "author_flair_css_class" key, and the corresponding value represents a unique flair this comment uses. The value is a string with the team's name and the flair id. For example, the flair values of the Log Angeles Lakers are phrased as "Lakers1", "Lakers2", "Lakers3", and "Lakers4".

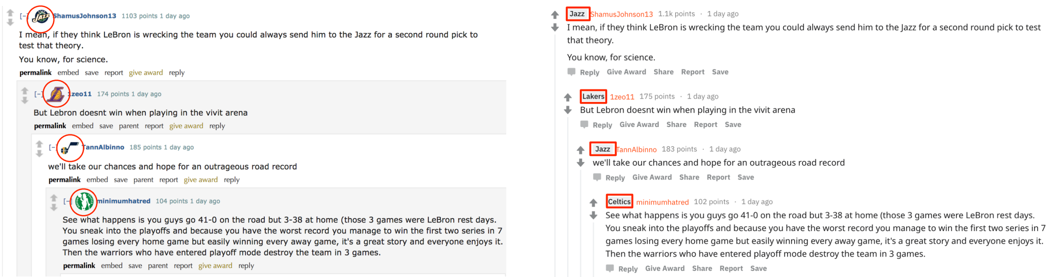


Fig. 3. An example of the flair usage in /r/NBA before and after the design change. Before the design change, flairs are represented by team logos while after the change, flairs are represented by the team name.

Table 1. The number of intergroup and single-group members in the 2018, 2017, and 2016 seasons.

	2018	2017	2016
Single-group	6,023	5,941	4,843
Intergroup	28,296	24,528	20,467

season. In other words, all activities of a fan indicate support towards his/her affiliated team. It follows that not every active user in NBA-related subreddits is identified as a fan of some team.

We further determine whether a fan of a team is exposed to intergroup contact based on his/her (lack of) behavior in /r/NBA, which we refer to as intergroup status. To summarize, we categorize fans of a team into the following two categories:

- **Intergroup**: Fans of a team who posted in both the affiliated team subreddit and /r/NBA in the season.
- **Single-group**: Fans of a team who had no activity in /r/NBA throughout the season.

Table 1 shows the number of members in each category. Since our study is concerned with intragroup behavior, i.e., behavior in the affiliated team subreddit, we view these intergroup and single-group fans as intergroup and single-group members of the affiliated team and study their behavior in the affiliated team subreddit.

Figure 4 presents the number of intergroup and single-group members in all 30 team subreddits in the 2018 season (see Figure A6 for the numbers in the 2017 and 2016 seasons). In every team subreddit, there are many more intergroup members than single-group members. Our definitions are based on user behavior in a single NBA season, and the label of a user can change across seasons. However, a single-group member rarely becomes an intergroup member in the next season in our dataset (6.0% of single-group members become intergroup members from 2016 to 2017, and 8.5% of single-group members become intergroup members from 2017 to 2018). This also confirms the tendency of single-group members to avoid intergroup contact.

We use both posting and commenting behavior to identify fans' team affiliation, but we focus on analyzing comments in the rest of the paper since the posts are usually much longer and more formal, and are thus not comparable to comments.

4 METHODS

We study the behavioral differences between intergroup and single-group members in their affiliated team subreddit by examining the expression of emotions in their comments in team subreddits for

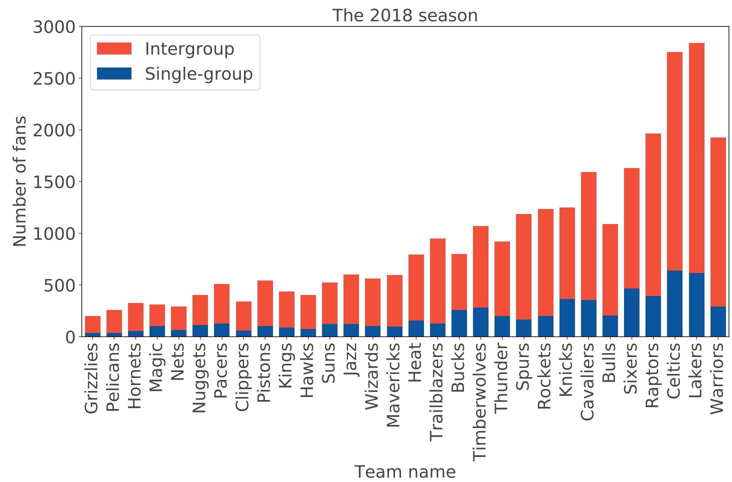


Fig. 4. The number of intergroup and single-group members affiliated with each NBA team in the 2018 season. We rank 30 team subreddits by the number of subscribers each team has by the end of the 2018 season. (see Figure A6 for the data statistics in the 2017 and 2016 seasons).

two reasons: (1) emotion is a central theme in understanding sports fans and their opinions; (2) textual content constitutes the main observed behavior in NBA-related discussion forums [53, 97, 113].

4.1 RQ1: Intragroup Behavior Differences

Matched intergroup members. A naïve way to compare these two categories of users is to directly examine all users in each category. However, such an approach does not take into account other important confounding factors, such as how active a member is in the group. We thus seek to ensure that intergroup and single-group members are a priori balanced on any observable features in the affiliated team subreddit, which indicates similar loyalty to the team. To achieve this, we adopt matching techniques: for each single-group member, we match him/her with the most similar unmatched intergroup member from the same affiliated team, where similarity is based on all the observed features. Due to the observational nature, whether a member has intergroup contact or not is not randomly assigned. In other words, our study reflects the behavioral differences between those who engage in intergroup contact and those who do not.

Following prior studies on factors associated with fan behavior in online sports communities [54, 57, 91, 113], we consider the following observable feature set for matching: (1) the number of comments in the affiliated team subreddit, (2) the average time gap between comments, (3) the average length of comments, (4) the proportion of comments in the playoff season, and (5) the proportion of comments in the game threads (these game threads are created for discussions during a game). All the comments examined here are in the members' affiliated team subreddit. We collect all of these feature values for each season. The similarities between fans are estimated using the nearest neighbor matching technique [95]. Min-max normalization is applied to each feature before feeding it into the matching model so that no single feature dominates the matching. We do not include the feedback (upvotes/downvotes) that members received from the NBA subreddits for matching because it can be endogenous with the language used in the comments (e.g., comments with hate speeches may not get many upvotes).

To evaluate the outcome of our matching procedure, for each observable feature, we check distributional differences between the treatment group (intergroup members) and the control group (single-group members). We compare their empirical cumulative distributions before and after matching using the Mann-Whitney U test [56]. The results of the 2018, 2017, and 2016 seasons are summarized in Figure A7. A small p-value here indicates that there exists a significant difference between the treatment group and the control group. Prior to matching, the p-value for each feature is close to 0, implying that the distributions do differ between groups. After matching, we find no difference between the treatment group and the matched control group for any observable feature at the 5% significance level ($\alpha = 0.05$) in all three seasons, indicating that the data is balanced across all the covariates after matching.

Language usage analysis. The proportion of emotional words (i.e., positive emotions, negative emotions, and swear words) in members' comments are analyzed using the Linguistic Inquiry and Word Count software (LIWC [77]), a word frequency-based text analysis tool (see Table A1 for examples of emotional words detected using this software). The hate speech comments are identified using an automated hate speech detection model [23]. It is a multi-class classifier that can reliably separate hate speech from other offensive language (see Table A2 for examples of hate speech comments detected using this detection model). According to Davidson et al. [23], the model achieved an overall precision of 0.91, recall of 0.90, and F1 score of 0.90 on detecting hate speech tweets.

Fightin-Words model. To identify a list of distinguishing keywords that are over-used by intergroup or single-group members, we apply the Fightin-Words algorithm [66] to compare the word frequencies to the background frequencies found in the other fan group's corpora using the informative Dirichlet prior model. This method estimates the log-odds ratio of each word w between two corpora α and β given the frequencies obtained from the background corpus $\mathcal{D}$. Then the log-odds ratio $\delta_w^{(\alpha\text{-}\beta)}$ for word w can be estimated as:

$$\delta_w^{(\alpha\text{-}\beta)} = log\frac{c_w^\alpha + c_w^\mathcal{D}}{c^\alpha + c^\mathcal{D} - c_w^\alpha + c_w^\mathcal{D}} - log\frac{c_w^\beta + c_w^\mathcal{D}}{c^\beta + c^\mathcal{D} - c_w^\beta + c_w^\mathcal{D}}, \tag{1}$$

where c_w^α and c_w^β are the counts of word w in corpora α and β, c^α and c^β are the counts of all words in corpora α and β, $c_w^\mathcal{D}$ is the count of word w in the background corpus $\mathcal{D}$, and $c^\mathcal{D}$ is the count of all words in corpus $\mathcal{D}$. The Fightin-Words algorithm also provides an estimation for the variance of the log-odds ratio,

$$\sigma^2(\delta_w^{(\alpha\text{-}\beta)}) \backsim \frac{1}{c_w^\alpha + c_w^\mathcal{D}} + \frac{1}{c_w^\beta + c_w^\mathcal{D}}, \tag{2}$$

and the corresponding z-score can be calculated as follows:

$$Z = \frac{\delta_w^{(\alpha\text{-}\beta)}}{\sqrt{\sigma^2(\delta_w^{(\alpha\text{-}\beta)})}}. \tag{3}$$

The Fightin-Words model is known to outperform other traditional methods in detecting word usage differences between corpora, such as PMI (pointwise mutual information) [58] and TF-IDF [87], by not over-emphasizing fluctuations of rare words [66]. We use the comments made by intergroup members as the background corpus for single-group members and vice versa to identify differences in language usage by each fan group [3]. We rank each word by averaging its z-scores calculated by the Fightin-Words model across all 30 teams. A higher positive z-score indicates this

[3]Part of our code is borrowed from Jack Hessel's Fightin-Words model implementation [43].

word is over-used by single-group members, and a higher negative z-score means this word is over-used by intergroup members.

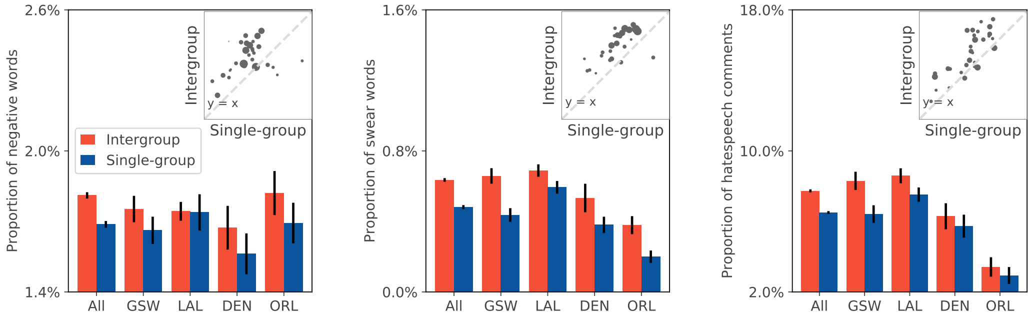
4.2 RQ2: Different Levels of Intergroup Contact

Matched intergroup members with different levels of intergroup contact. We define a user's level of intergroup contact based on the fraction of comments in the intergroup setting (/r/NBA). The fraction is calculated as the proportion of the number of comments the user made in the /r/NBA versus the total number of comments in NBA-related subreddits. Specifically, for each single-group member, we again apply the nearest neighbor matching technique to find five closest intergroup members in the same affiliated team and assign a label of 1, 2, 3, 4, or 5 to them based on their fraction of comments in /r/NBA in a complete NBA season. Different from pairing intergroup members and single-group members before, we do it with replacement because there are not enough intergroup members to conduct this matching uniquely. As such, an intergroup member can be matched to multiple single-group members. We compare the empirical cumulative distributions before and after matching for each level using the Mann-Whitney U test [35]. The results of the 2018, 2017, and 2016 seasons are presented in Figure A8, A9, and A10, respectively. We also assign a label of 0 to single-group members. A larger label indicates a higher level of intergroup contact that the member has in /r/NBA. We aggregate intergroup members at each level across all 30 team subreddits to compare their intragroup behavior. Note that the number of members at each level is the same, but some intergroup members may be counted more than once.

Regression analyses of the relationship between different levels of intergroup contact and language usage. To understand the relationship between members' intergroup contact level and language usage, we also conduct OLS regression analyses after the above matching procedure. The independent variables considered in the regression model are the same set of features used in matching intergroup and single-group members (Section 4.1). We standardize all independent variables before feeding into the regression model. Our full linear regression model to test each language usage pattern is shown below:

$$\begin{aligned}
\textit{Proportion of language usage} \sim & \beta_0 + \beta_1 \textit{number of comments} + \beta_2 \textit{average comment hours gap} \\
& + \beta_3 \textit{average comment length} + \beta_4 \textit{proportion of playoff comments} \\
& + \beta_5 \textit{proportion of game thread comments} \\
& + \beta_6 \textit{fraction} + \beta_7 \textit{fraction} \times \textit{level1} + \beta_8 \textit{fraction} \times \textit{level2} \\
& + \beta_9 \textit{fraction} \times \textit{level3} + \beta_{10} \textit{fraction} \times \textit{level4} + \beta_{11} \textit{fraction} \times \textit{level5}
\end{aligned}$$

The fraction in the linear regression model refers to the proportion of the number of comments the user made in the intergroup setting (/r/NBA) versus the total number of comments in NBA-related subreddits. All the control variables for matching intergroup members and single-group members are included. There are repeated measures in our regression model as an intergroup member can be matched to more than one single-group member. The average number of times for an intergroup member to be matched is 1.77 (excluding the intergroup members who never get matched). Among the intergroup members who are matched more than once, the average variance of their intergroup contact levels in different matches is 0.27. The small variance shows the consistency of our matching technique.



(a) Proportion of negative words.

(b) Proportion of swear words.

(c) Proportion of hate speech comments.

Fig. 5. The comparison of language usage between intergroup and single-group members in the 2018 season. Intergroup members use more negative words (Figure 5a; two-tailed t-test, $t = 6.23$, $p < 0.001$, 95% CI=0.08% to 0.16%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (Figure 5b; two-tailed t-test, $t = 3.51$, $p < 0.001$, 95% CI=0.02% to 0.08%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$), and generate more hate speech comments (Figure 5c; two-tailed t-test, $t = 10.44$, $p < 0.001$, 95% CI=1.00% to 1.46%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$). "All" is based on concatenating the samples from all 30 NBA team subreddits, and we also show the top two and bottom two teams ranked by the number of subscribers that have at least 100 single-group members. We further show the scatter plot of all 30 teams in the top right to illustrate that the findings are robust across teams (the size of the dot is proportional to the number of subscribers). Error bars represent standard errors. The results are consistent in the 2017 and 2016 seasons (see Figure A11 and Figure A12).

5 RESULTS

In this section, we examine intragroup language differences between intergroup and single-group members (RQ1). We further discuss how different levels of intergroup contact relate to intragroup behavior (RQ2).

5.1 RQ1: Intragroup Language Differences

Figure 5 compares negative language usage between matched intergroup and single-group members. Intergroup members tend to use more negative language than single-group members, which is indicated by the use of more negative words (two-tailed t-test, $t = 6.23$, $p < 0.001$, 95% CI=0.08% to 0.16%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (two-tailed t-test, $t = 3.51$, $p < 0.001$, 95% CI=0.02% to 0.08%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) based on lexicon analysis. We further compute the proportion of hate speech with an automated hate speech and offensive language detection model [23]. It is consistent that intergroup members also generate more hate speech (two-tailed t-test, $t = 10.44$, $p < 0.001$, 95% CI=1.00% to 1.46%; 26 out of 30 teams, two-tailed binomial test $p < 0.001$). These results indicate that intergroup members are more emotionally charged in their intragroup behavior compared with the single-group members and are somewhat different from the hypothesis that intergroup contact enhances empathy and perspective thinking [4, 80, 94]. Our results are consistent when excluding the NBA playoffs (see Section A.1). We also compare positive language usage between intergroup members and single-group members and do not find a consistent trend at the 5% significance level (see Figure A13).

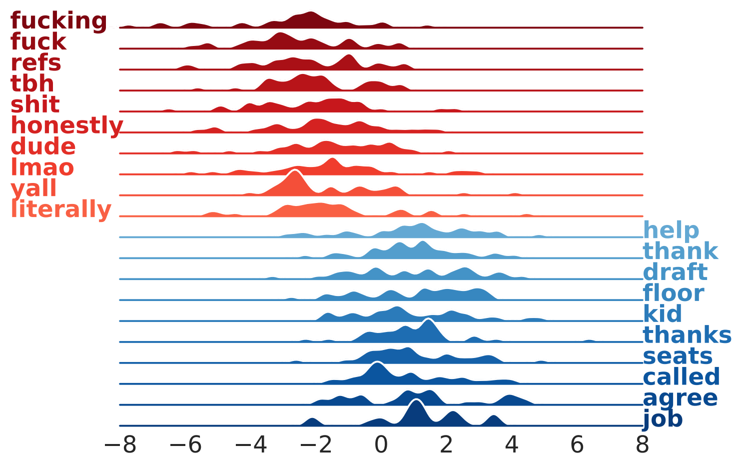


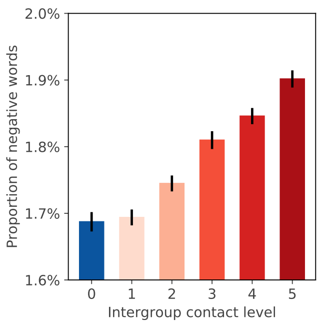
Fig. 6. The top-10 over-represented words used by intergroup (red) and single-group (blue) members in the 2018 season. For each word, we show the distribution of the z-scores for all 30 teams calculated by the Fightin-Words algorithm [66].

To further understand the difference between intergroup and single-group members in language usage, we identify a list of distinguishing words that are more likely to be used by intergroup or by single-group members, using the Fightin-Words algorithm with the informative Dirichlet prior model [66]. Figure 6 lists the top-10 over-represented words used by intergroup and single-group members in the 2018 season. We rank each word by its average z-score calculated by the Fightin-Words algorithm across all 30 teams. A positive z-score indicates that this word is over-used by single-group members, while a negative z-score suggests that this word is over-used by intergroup members. Our results show that single-group members are more friendly and calm when commenting in the affiliated team subreddit and use more polite words, such as “agree”, “thanks”, and “help”. Also, “seats” suggest that some single-group members are local fans, as they frequently discuss information about attending live games. In comparison, intergroup members use more swear words and talk more about the referees (likely complaining).

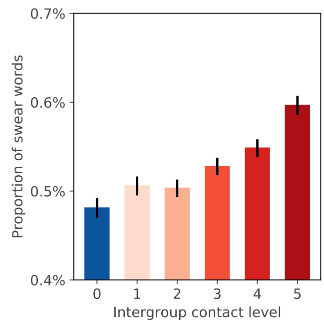
5.2 RQ2: Different Levels of Intergroup Contact

In addition to identifying the intragroup behavioral differences, our observational study allows us to quantify different levels of intergroup contact, which can be difficult to operationalize in experimental studies. Here, we examine the mechanisms of how increased levels of intergroup contact relate to differences in intragroup behavior.

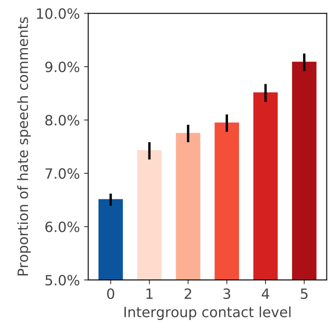
Figure 7 shows language usage differences between members with different intergroup contact levels. Members of higher intergroup contact levels are generally more negative in language usage: They tend to use more negative words (mean = 1.69%, 1.69%, 1.76%, 1.80%, 1.85%, and 1.90%, respectively for labels from 0 to 6) and swear words (mean = 0.48%, 0.50%, 0.50%, 0.53%, 0.55%, and 0.59%, respectively for labels from 0 to 6), and generate more hate speech comments (mean = 6.51%, 7.61%, 7.72%, 7.77%, 8.43%, and 9.15%, respectively for labels from 0 to 6) in the affiliated team subreddit. However, the trends are not necessarily linear. For instance, intergroup members at level 1 do not show significant differences from single-group members in negative word usage,



(a) Proportion of negative words.



(b) Proportion of swear words.



(c) Proportion of hate speech comments.

Fig. 7. Intragroup language usage differences of members with different intergroup contact levels in the 2018 season. x-axis represents intergroup levels determined by the fraction of comments in /r/NBA. We observe a consistent monotonic pattern in the proportion of negative words (mean = 1.69%, 1.69%, 1.74%, 1.81%, 1.85%, and 1.90%, respectively for labels from 0 to 6;), swear words (mean = 0.48%, 0.51%, 0.50%, 0.53%, 0.55%, and 0.60%, respectively for labels from 0 to 6), and hate speech comments (mean = 6.51%, 7.42%, 7.75%, 7.94%, 8.51%, and 9.08%, respectively for labels from 0 to 6). The monotonic trend is consistent in the 2017 and 2016 season (see Figure A14 and Figure A15). Error bars represent standard errors.

while intergroup members at level 5 present a significant jump from previous levels in negative words, swear words, and the use of hate speech.

Table 2 shows the results of regression analyses. The fraction of intergroup contact has a statistically significant positive coefficient in regressions for the proportion of negative words, swear words, and hate speech comments. Moreover, the coefficients for some interaction terms with levels are also statistically significant (e.g., level 5, β_{11}, is statistically significant in regressions for the proportion of negative words, swear words, and hate speech comments), indicating that nonlinear corrections are required. Note that the BIC score is consistently better by incorporating the interaction terms, although adjusted R^2 remains the same due to the fact that this is a very challenging regression task.

6 INTRAGROUP BEHAVIOR VS. INTERGROUP BEHAVIOR OF THE SAME USER

Given the clear intragroup behavioral differences in language usage between intergroup and single-group members, we end our study by exploring the potential reasons behind them. We study the differences in language usage of the same user in his/her affiliated team subreddit vs. in /r/NBA. We compare the same person in two different contexts and naturally control for most of the confounding factors, which is also connected with the personality vs. situation debate [48].

Figure 8 shows that intergroup members use even more negative language in the intergroup setting, as they use more negative words (two-tailed t-test, $t = 9.39$, $p < 0.001$, 95% CI=0.17% to 0.26%; 30 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 13.69$, $p < 0.001$, 95% CI=0.22% to 0.29%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$), and generate more hate speech comments than in the intragroup setting (two-tailed t-test, $t = 2.97$, $p = 0.003$, 95% CI=0.18% to 0.88%; 23 out of 30 teams, two-tailed binomial test $p = 0.005$). This indicates that fans are more hostile when facing fans from other teams than from the same team. This observation is robust after controlling for topics of discussion by only considering game threads (see Section A.3 and Figure A5). Our observation suggests that although

Table 2. Regression analyses for the proportion of negative words, swear words, and hate speech comments. The variables used for matching intergroup and single-group members are also included for control. For each of the analyses, the fraction of intergroup contact has a positive coefficient. The number of stars indicates p-values, ***: $p < 0.001$, **: $p < 0.01$ *: $p < 0.05$.

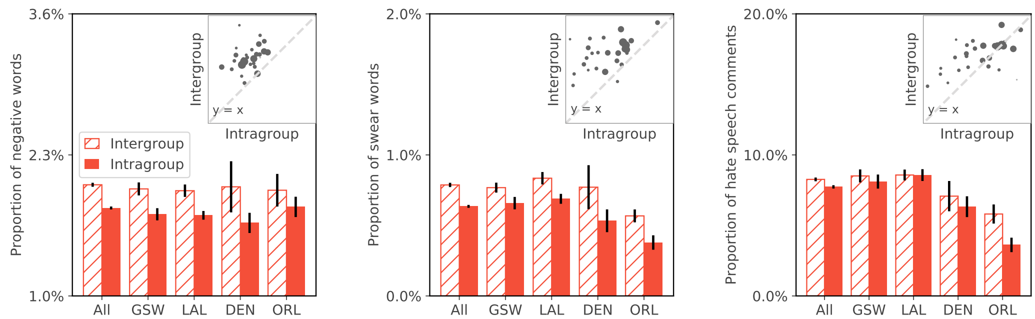
Variable	Prop. of negative words Reg. 1	Prop. of negative words Reg. 2	Prop. of swear words Reg. 1	Prop. of swear words Reg. 2	Prop. of hate speech Reg. 1	Prop. of hate speech Reg. 2
Control						
number of comments	0.003*	0.003*	0.023***	0.022***	0.002*	0.002*
average comment hours gap	−0.004***	0.004***	−0.017***	−0.017***	−0.003***	−0.003***
average comment length	−0.028***	−0.028***	−0.000	−0.000	−0.042***	−0.042***
Prop. of playoff comments	0.006***	0.006***	0.019***	0.018***	0.004***	0.004***
Prop. of game thread comments	0.053***	0.053***	0.090***	0.089***	0.028***	0.028***
Fraction						
fraction	0.007***	0.005***	0.027***	0.019***	0.004***	0.003***
Levels						
fraction × level1		−0.002		−0.001		0.001
fraction × level2		0.001		−0.001		−0.001
fraction × level3		0.002**		0.003		0.000
fraction × level4		0.002**		0.005*		0.001
fraction × level5		0.002***		0.010***		0.002***
intercept	0.052***	0.053***	0.040***	0.042***	0.028***	0.028***
Adjusted R^2	0.172	0.172	0.033	0.033	0.138	0.138
BIC	-147675	-147641	-45838	-45803	-164859	-164835

intergroup members are more emotional than single-group members in the affiliated subreddit, they are not as "outrageous" as they are in the intergroup setting. In comparison, when going to the intergroup setting and confronting fans from other team groups, they tend to have more negative interactions and troll each other.

These observations may provide explanations for the characteristics of intergroup fans in intragroup behavior. Prior studies suggest that negative intergroup contact is more influential in shaping people's attitudes and may curb the contact's ability to reduce prejudice [33, 74, 93]. The emotionally charged intergroup contact from the intergroup setting may connect to intergroup fans' more sentimental attitudes in their affiliated team subreddit. It requires further research to establish the causal link here, but the fact that we are able to observe these contrasts demonstrates the importance of such observational studies based on real interactions over substantial time periods.

7 DISCUSSION

Although most previous studies have focused on the role of intergroup contact in changing attitudes of individual members, our study highlights the fact that users selectively become intergroup members, and intergroup and single-group members in turn interact with each other in their affiliated group. Such interaction can potentially influence members' language usage and shape the entire group. Moreover, we demonstrate a variety of ways in which intergroup contact levels can moderate intragroup behavior. This indicates that observational studies can provide important complementary evidence to experimental studies on this topic because interventions can hardly



(a) Proportion of negative words.

(b) Proportion of swear words.

(c) Proportion of hate speech comments.

Fig. 8. Intergroup members use more negative language in the intergroup setting than in the intragroup setting in the 2018 season. Here we only consider the matched intergroup members in Figure 5 (i.e., the solid red bars in this figure are identical to the red bars in Figure 5). They use more negative words (two-tailed t-test, $t = 9.39$, $p < 0.001$, 95% CI=0.17% to 0.26%; 30 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 13.69$, $p < 0.001$, 95% CI=0.22% to 0.29%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$), and generate more hate speech comments (two-tailed t-test, $t = 2.97$, $p = 0.003$, 95% CI=0.18% to 0.88%; 23 out of 30 teams, two-tailed binomial test $p = 0.005$). "All" is based on concatenating the samples from all 30 NBA team subreddits, and we also show the top two and bottom two teams ranked by the number of subscribers that have at least 100 single-group members. We further show the scatter plot of all 30 teams in the top right to illustrate that the findings are robust across teams (the size of the dot is proportional to the number of subscribers). Error bars represent standard errors. The results are consistent in the 2017 and 2016 seasons (see Figure A16 and Figure A17).

result in deep and regular contact. Novel methodologies are required to further bridge the gap between observational studies and experimental studies.

Could social media be driving polarization? Twitter, Reddit, and Facebook have become important platforms for political discussions as well as misinformation [36, 104]. Service providers are designing new features that would actively expose people to opposing views. For example, Twitter recently experimented with new algorithms that would promote alternative viewpoints in Twitter's timeline to address misinformation and reduce the effect of echo chambers [86]. However, the proposed solution may increase polarization. Unlike decades of offline experiments which mostly indicate intimate contact between members of rival groups across an extended period can produce positive effects, the results in Bail et al. [3] and our paper suggest that encountering views from opposing groups online may make them even more wedded to their own views. There are several possible explanations of this contrast by examining the possible mechanisms that intergroup contact affects individual attitudes. First, the comments created on social media are usually brief. These short messages without enough context may not enhance knowledge about opposing groups. Several studies suggest that people interpret short text-based messages inconsistently, which creates significant potential for miscommunication [47, 62, 110]. Second, the discussion structure may facilitate the spread of negative interaction. Cheng et al. [13] examines the evolution of discussions on CNN.com and show that existing trolling comments in a discussion thread significantly increase the likelihood of future trolling comments. The spread of negativity will increase rather than reduce people's anxiety levels when facing opposing groups. Third, the anonymous, spontaneous nature of communications on social media may not be conducive to cultivating empathy. In an experiment designed to examine the relationship between the presence of mobile devices and the quality of social interactions, results show that participants who have conversations in the absence of mobile devices report high levels of empathetic concern [65]. In summary, intergroup contact may lead to diverging outcomes depending on the environment and the nature of the contact. Further research is required to examine these possibilities and understand how social and technical design decisions can influence the outcomes.

Can we design better online discussion forums for different groups? The findings in this work indicate that social platforms designers should consider strategies to shape intergroup contact online. As hinted above, it is insufficient to recommend users to follow members of opposing groups or opposing views. Better design strategies need to be experimented for encouraging civil and extended intergroup contact. It would also be useful to take into account how different levels of intergroup contact may moderate individual opinions differently. Content moderation can be a promising area for future studies in the context of intergroup contact [49]. For instance, Matias [61] shows the displaying community rules can prevent harassment, but how to reduce negative intergroup contact remains an open question. Similarly, a powerful way of spreading online information is through social consensus cues and online endorsement (e.g., upvotes, likes). However, promoting content with the highest popularity can sometimes be problematic. Earlier research suggests that tweets with more sentiment-laden words are likely to be favorited or retweeted, and politicians may intentionally use this strategy to maximize impacts on Twitter [9, 99, 101]. Our study also finds that intergroup members receive better feedback from their affiliated team subreddit even though they use more negative language (Figure A4). This type of behavior can generate negative reactions from opposing groups and push the whole discussion to cycle towards more emotionally-laden and potentially polarizing content. It is thus important to develop comment ranking systems that are cognizant of intergroup contact and prioritizes constructive interactions.

Limitations. Our findings are subject to the following limitations. First, the causal relationship between intergroup contact and negative language usage is not entirely clear. Due to the nature of our observational study, whether a member has intergroup contact is not randomly assigned. Though we match users based on a series of activity features, an important confounding factor could be that people who seek intergroup contact are inherently different from those who do not.

Second, our definition of intergroup contact entails that we focus on relatively active users. Thus, we cannot observe indirect intergroup contact, such as browsing /r/NBA. Prior studies have shown that indirect contact, such as imagining oneself interacting with an out-group member and observing an in-group member interacting with an out-group member [27, 103], may also shape human behavior. It also follows that intergroup members have more activities on NBA-related discussion forums as a whole than single-group members. We want to note that the nature of intergroup contact is that given the same amount of time in life, individuals with intergroup contact put more effort into intergroup contact than those without such contact.

Third, we use a coarse proxy to consider any users who have posted in our intergroup setting (/r/NBA) as intergroup members, and study the language differences in the intragroup setting at the user level instead of at the dyad level. However, some comments created by the fans in /r/NBA may be replies to the fans who are from the same team or do not have a team affiliation. More in-depth characterization of different types of discussions happened in the intergroup setting is required to further understand the differences observed in this study.

Fourth, the observations made in this study are limited to Reddit NBA fan groups. The sports context might be a strong case for understanding intergroup relations, as all the teams are created to compete with each other for the final championship. The expression of hostile attitudes towards opposing sides are culturally acceptable and even encouraged [17]. We should expect less negative intergroup contact between groups that do not contend for the same resources (e.g., music fans of different musicians may not have conflicts with each other at all). However, politics, especially in a polarized bipartisan situation, share common properties with the sports context. Examining the generalization of our results in other contexts is a promising avenue for future work.

Finally, the negative language observed in our study may not necessarily bring negative effects to the community. Prior studies suggest the main reason people use swearing words on the online platform is to express some strong emotions, such as anger and frustration [13, 25, 105]. Heath et al. [42] examine users' emotional selection in memes when emotion is manipulated and observe that people prefer the version of the story that produced the highest levels of disgust and evoke strong sentiment. Jay [46] further argues that only when cursing occurs in the form of insults toward others, such as name-calling, harassment, and hate speech, it becomes harmful. In addition, earlier literature suggests that the reason people use swearing words on the online platform may relate to Internet humor, such as jokes and memes. Posting humorous content on the Internet has the potential to engage other users in art activities that are closely connected to their lives and receive online endorsements [90, 111]. Attempting to be funny could be another reason that intergroup members adopt a more negative language style than single-group members. However, as pointed out by Lockyer and Pickering [55], a significant proportion of Internet humor has offensive, sexism, and racism content, and its consequences are often overlooked.

8 CONCLUSION

In this paper, by applying our computational framework to NBA-related discussion forums on Reddit, we identify clear language differences between intergroup and single-group members in their affiliated group (the intragroup setting). We find that in the affiliated team subreddit, intergroup members tend to use more negative and swear words, and generate more hate speech comments compared with single-group members. Moreover, we quantify different levels of intergroup contact for each intergroup member based on the fraction of their comments in the intergroup setting (/r/NBA). Interestingly, the level of intergroup contact can relate to differences in language usage in different ways, though the relationship is mostly monotonic. To further shed light on the behavior of intergroup members, we also compare the language usage of intergroup members between the intragroup setting and the intergroup setting. This setup naturally controls for the subject because we compare the same person across two different environments. We observe that intergroup members are even more negative and more likely to swear in the intergroup setting. As intergroup contact in online platforms becomes increasingly common and can play an important role in opinion formation, our work demonstrates how observational studies can provide complementary evidence to experimental studies on this topic.

REFERENCES

[1] Nicola Abbott and Lindsey Cameron. 2014. What makes a young assertive bystander? The effect of intergroup contact, empathy, cultural openness, and in-group bias on assertive bystander intervention intentions. *Journal of Social Issues* 70, 1 (2014), 167–182.

[2] Christopher A Bail. 2014. *Terrified: How anti-Muslim fringe organizations became mainstream.* Princeton University Press.

[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* (2018).

[4] C Daniel Batson, Marina P Polycarpou, Eddie Harmon-Jones, Heidi J Imhoff, Erin C Mitchener, Lori L Bednar, Tricia R Klein, and Lori Highberger. 1997. Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology* 72, 1 (1997), 105.

[5] Jason Baumgartner. 2018. Reddit Dataset. https://files.pushshift.io/reddit/. [Online; accessed 15-June-2019].

[6] Daniel J Beal, Janet B Ruscher, and Sherry B Schnake. 2001. No benefit of the doubt: Intergroup bias in understanding causal explanation. *British Journal of Social Psychology* 40, 4 (2001), 531–543.

[7] Jeremy Blackburn and Haewoon Kwak. 2014. Stfu noob!: predicting crowdsourced decisions on toxic behavior in online games. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 877–888.

[8] Johanne Boisjoly, Greg J Duncan, Michael Kremer, Dan M Levy, and Jacque Eccles. 2006. Empathy or antipathy? The impact of diversity. *American Economic Review* 96, 5 (2006), 1890–1905.

[9] William J Brady, Julian A Wills, Dominic Burkart, John T Jost, and Jay J Van Bavel. 2018. An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General* (2018).

[10] Braz Camargo, Ralph Stinebrickner, and Todd Stinebrickner. 2010. Interracial friendships in college. *Journal of Labor Economics* 28, 4 (2010), 861–892.

[11] Ellis Cashmore. 2005. *Making sense of sports*. Routledge.

[12] 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), 31.

[13] 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*. ACM, 1217–1230.

[14] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2014. How community feedback shapes user behavior. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

[15] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial behavior in online discussion communities. In *Proceedings of International AAAI Conference on Web and Social Media*.

[16] Amy Chua. 2018. *Political Tribes: Group Instinct and the Fate of Nations*. Penguin.

[17] Mina Cikara, Matthew M Botvinick, and Susan T Fiske. 2011. Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological Science* 22, 3 (2011), 306–313.

[18] Mina Cikara and Susan T Fiske. 2013. Their pain, our pleasure: stereotype content and schadenfreude. *Annals of the New York Academy of Sciences* 1299, 1 (2013), 52–59.

[19] Marci D Cottingham. 2012. Interaction ritual theory and sports fans: Emotion, symbols, and solidarity. *Sociology of Sport Journal* 29, 2 (2012), 168–185.

[20] Bernard Cova and Stefano Pace. 2006. Brand community of convenience products: new forms of customer empowerment–the case "my Nutella The Community". *European Journal of Marketing* 40, 9/10 (2006), 1087–1105.

[21] Vassilis Dalakas and Aron M Levin. 2005. The balance theory domino: How sponsorships may elicit negative consumer attitudes. *ACR North American Advances* (2005).

[22] Srayan Datta and Eytan Adar. 2019. Extracting Inter-Community Conflicts in Reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 146–157.

[23] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. In *Proceedings of International AAAI Conference on Web and Social Media*. 512–515.

[24] Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. 2014. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 626–638.

[25] Nicholas Diakopoulos and Mor Naaman. 2011. Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*. ACM, 133–142.

[26] Bradley Doucet. 2010. Why Are Sports Fans so Biased? https://goo.gl/mQZTXQ. [Online; accessed 15-June-2019].

[27] John F Dovidio, Anja Eller, and Miles Hewstone. 2011. Improving intergroup relations through direct, extended and other forms of indirect contact. *Group Processes & Intergroup Relations* 14, 2 (2011), 147–160.

[28] John F Dovidio, Samuel L Gaertner, and Kerry Kawakami. 2003. Intergroup contact: The past, present, and the future. *Group Processes & Intergroup Relations* 6, 1 (2003), 5–21.

[29] John F Dovidio, Angelika Love, Fabian MH Schellhaas, and Miles Hewstone. 2017. Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. *Group Processes & Intergroup Relations* 20, 5 (2017), 606–620.

[30] Steve Frosdick and Peter Marsh. 2013. *Football Hooliganism*. Willan.

[31] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. 2018. Me, my echo chamber, and I: introspection on social media polarization. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 823–831.

[32] Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 253–262.

[33] Sylvie Graf, Stefania Paolini, and Mark Rubin. 2014. Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five Central European countries. *European Journal of Social Psychology* 44, 6 (2014), 536–547.

[34] Donald P Green and Janelle S Wong. 2009. Tolerance and the contact hypothesis: A field experiment. *The Political Psychology of Democratic Citizenship* (2009), 1–23.

[35] 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*. ACM, 1400–1408.

[36] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on twitter during the 2016 US Presidential election. *Science* 363, 6425 (2019), 374–378.

[37] Anatoliy Gruzd and Jeffrey Roy. 2014. Investigating political polarization on Twitter: A Canadian perspective. *Policy & Internet* 6, 1 (2014), 28–45.

[38] Pedro Henrique Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A Measure of Polarization on Social Media Networks Based on Community Boundaries.. In *Proceedings of International AAAI Conference on Web and Social Media*.

[39] Richard Guilianotti. 2013. *Football, violence and social identity*. Routledge.

[40] Allen Guttmann. 2004. *From ritual to record: The nature of modern sports*. Columbia University Press.

[41] Cody T Havard. 2014. Glory out of reflected failure: The examination of how rivalry affects sport fans. *Sport Management Review* 17, 3 (2014), 243–253.

[42] Chip Heath, Chris Bell, and Emily Sternberg. 2001. Emotional selection in memes: the case of urban legends. *Journal of Personality and Social Psychology* 81, 6 (2001), 1028.

[43] Jack Hessel. 2016. Fighting Words Model Quick Implementation. https://github.com/jmhessel/FightingWords. [Online; accessed 15-June-2019].

[44] Thomas Hickman and James Ward. 2007. The dark side of brand community: Inter-group stereotyping, trash talk, and schadenfreude. *ACR North American Advances* (2007).

[45] Kathleen Hall Jamieson and Joseph N Cappella. 2008. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press.

[46] Timothy Jay. 2009. Do offensive words harm people? *Psychology, Public Policy, and Law* 15, 2 (2009), 81.

[47] Lynne Kelly, James A Keaten, Bonnie Becker, Jodi Cole, Lea Littleford, and Barrett Rothe. 2012. ŞIt's the American lifestyle!Ť: An investigation of text messaging by college students. *Qualitative Research Reports in Communication* 13, 1 (2012), 1–9.

[48] Douglas T Kenrick and David C Funder. 1988. Profiting from controversy: Lessons from the person-situation debate. *American Psychologist* 43, 1 (1988), 23.

[49] Sara Kiesler, Robert Kraut, Paul Resnick, and Aniket Kittur. 2012. Regulating behavior in online communities. *Building Successful Online Communities: Evidence-Based Social Design. MIT Press, Cambridge, MA* (2012).

[50] Nuri Kim and Magdalena Wojcieszak. 2018. Intergroup contact through online comments: Effects of direct and extended contact on outgroup attitudes. *Computers in Human Behavior* 81 (2018), 63–72.

[51] Srijan Kumar, William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2018. Community interaction and conflict on the web. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 933–943.

[52] Changjun Lee, Jieun Shin, and Ahreum Hong. 2018. Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea. *Telematics and Informatics* 35, 1 (2018), 245–254.

[53] Steven A Lehr, Meghan L Ferreira, and Mahzarin R Banaji. 2019. When outgroup negativity trumps ingroup positivity: Fans of the Boston Red Sox and New York Yankees place greater value on rival losses than own-team gains. *Group Processes & Intergroup Relations* 22, 1 (2019), 26–42.

[54] Weiwen Leung, Haiyi Zhu, and Joseph A. Konstan. 2017. The Effect of Emotional Cues from the NFL on Wikipedia Contributions. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 66 (2017), 66:1–66:21 pages.

[55] Sharon Lockyer and Michael Pickering. 2005. *Beyond a joke: The limits of humour*. Springer.

[56] Henry B Mann and Donald R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* (1947), 50–60.

[57] Leon Mann. 1989. Sports crowds and the collective behavior perspective. *Sports, games, and play: Social and psychological viewpoints* (1989), 299–331.

[58] Christopher D Manning, Christopher D Manning, and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.

[59] Dejan Markovikj, Sonja Gievska, Michal Kosinski, and David J Stillwell. 2013. Mining facebook data for predictive personality modeling. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

[60] Lilliana Mason. 2018. *Uncivil agreement: How politics became our identity*. University of Chicago Press.

[61] J Nathan Matias. 2019. Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. *Proceedings of the National Academy of Sciences* 116, 20 (2019), 9785–9789.

[62] Hannah Miller, Daniel Kluver, Jacob Thebault-Spieker, Loren Terveen, and Brent Hecht. 2017. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. In *Proceedings of the International AAAI Conference on Web and Social Media*.

[63] Atsunori Minamikawa and Hiroyuki Yokoyama. 2011. Blog tells what kind of personality you have: egogram estimation from Japanese weblog. In *Proceedings of the ACM 2011 Conference on Computer-Supported Cooperative Work*. ACM, 217–220.

[64] Atsunori Minamikawa and Hiroyuki Yokoyama. 2011. Personality estimation based on weblog text classification. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 89–97.

[65] Shalini Misra, Lulu Cheng, Jamie Genevie, and Miao Yuan. 2016. The iPhone effect: the quality of in-person social interactions in the presence of mobile devices. *Environment and Behavior* 48, 2 (2016), 275–298.

[66] Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. 2008. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16, 4 (2008), 372–403.

[67] Albert M Muniz and Thomas C O'guinn. 2001. Brand community. *Journal of Consumer Research* 27, 4 (2001), 412–432.

[68] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai. 2010. Is it really about me?: message content in social awareness streams. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*. ACM, 189–192.

[69] NBA.com. 2018. NBA breaks all-time attendance record for fourth straight year. http://www.nba.com/article/2018/04/12/nba-breaks-attendance-record-fourth-straight-year. [Online; accessed 15-June-2019].

[70] Nielsen.com. 2015. The Year in Sports Media Report: 2015. https://www.nielsen.com/us/en/insights/reports/2016/the-year-in-sports-media-report-2015.html. [Online; accessed 15-June-2019].

[71] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[72] Erik L Olson. 2018. Are rival team fans a curse for home team sponsors? The moderating effects of fit, oppositional loyalty, and league sponsoring. *Marketing Letters* 29, 1 (2018), 115–122.

[73] Elizabeth Page-Gould, Rodolfo Mendoza-Denton, and Linda R Tropp. 2008. With a little help from my cross-group friend: Reducing anxiety in intergroup contexts through cross-group friendship. *Journal of Personality and Social Psychology* 95, 5 (2008), 1080.

[74] Stefania Paolini, Jake Harwood, and Mark Rubin. 2010. Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin* 36, 12 (2010), 1723–1738.

[75] Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2015. Automatic personality assessment through social media language. *Journal of Personality and Social Psychology* 108, 6 (2015), 934.

[76] Warren Pearce, Kim Holmberg, Iina Hellsten, and Brigitte Nerlich. 2014. Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report. *PloS one* 9, 4 (2014), e94785.

[77] James W Pennebaker, Martha E Francis, and Roger J Booth. 2007. Linguistic inquiry and word count: LIWC 2007. *Technical report* (2007).

[78] Thomas F Pettigrew. 1997. Generalized intergroup contact effects on prejudice. *Personality and Social Psychology Bulletin* 23, 2 (1997), 173–185.

[79] 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.

[80] Thomas F Pettigrew and Linda R Tropp. 2008. How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology* 38, 6 (2008), 922–934.

[81] Thomas F Pettigrew, Linda R Tropp, Ulrich Wagner, and Oliver Christ. 2011. Recent advances in intergroup contact theory. *International Journal of Intercultural Relations* 35, 3 (2011), 271–280.

[82] Daniele Quercia, Renaud Lambiotte, David Stillwell, Michal Kosinski, and Jon Crowcroft. 2012. The personality of popular facebook users. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*. ACM, 955–964.

[83] Reddit. 2018. Top Reddits. http://redditmetrics.com/top. [Online; accessed 15-June-2019].

[84] Redditlist. 2018. Most Active Subreddits. http://redditlist.com/all. [Online; accessed 15-June-2019].

[85] Alan Roadburg. 1980. Factors precipitating fan violence: A comparison of professional soccer in Britain and North America. *British Journal of Sociology* (1980), 265–276.

[86] Tony Romm and Dwoskin Elizabeth. 2018. Jack Dorsey says he's rethinking the core of how Twitter works. https://www.washingtonpost.com/technology/2018/08/15/jack-dorsey-says-hes-rethinking-core-how-twitter-works/?utm_term=.ca3c42463c6f. [Online; accessed 15-June-2019].

[87] Gerard Salton and Michael J McGill. 1986. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc.

[88] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.
[89] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping pro and anti-social behavior on Twitch through moderation and example-setting. In *Proceedings of Computer-Supported Cooperative Work and Social Computing (CSCW) '17*. ACM, New York, NY, USA. http://doi.acm.org/10.1145/2998181.2998277
[90] Limor Shifman. 2014. *Memes in digital culture*. Mit Press.
[91] Eric Simons. 2015. The psychology of why sports fans see their teams as extensions of themselves. https://www.washingtonpost.com/opinions/the-psychology-of-why-sports-fans-see-their-teams-as-extensions-of-themselves/2015/01/30/521e0464-a816-11e4-a06b-9df2002b86a0_story.html. [Online; accessed 15-June-2019].
[92] Marc A Smith, Lee Rainie, Ben Shneiderman, and Itai Himelboim. 2014. Mapping Twitter topic networks: From polarized crowds to community clusters. *Pew Research Center* 20 (2014), 1–56.
[93] Tobias H Stark, Andreas Flache, and René Veenstra. 2013. Generalization of positive and negative attitudes toward individuals to outgroup attitudes. *Personality and Social Psychology Bulletin* 39, 5 (2013), 608–622.
[94] Walter G Stephan and Krystina Finlay. 1999. The role of empathy in improving intergroup relations. *Journal of Social issues* 55, 4 (1999), 729–743.
[95] Elizabeth A Stuart. 2010. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* 25, 1 (2010), 1.
[96] Cass R Sunstein. 2009. *Going to extremes: How like minds unite and divide*. Oxford University Press.
[97] William A Sutton, Mark A McDonald, George R Milne, and John Cimperman. 1997. Creating and fostering fan identification in professional sports. *Sport Marketing Quarterly* 6 (1997), 15–22.
[98] Henri Tajfel. 1982. Social psychology of intergroup relations. *Annual Review of Psychology* 33, 1 (1982), 1–39.
[99] Chenhao Tan, Adrien Friggeri, and Lada Adamic. 2016. Lost in propagation? Unfolding news cycles from the source. In *Proceedings of ICWSM*.
[100] Chenhao Tan and Lillian Lee. 2015. All who wander: On the prevalence and characteristics of multi-community engagement. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1056–1066.
[101] Chenhao Tan, Lillian Lee, and Bo Pang. 2014. The effect of wording on message propagation: Topic-and author-controlled natural experiments on Twitter. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Vol. 1. 175–185.
[102] Catalina L Toma and Jeffrey T Hancock. 2010. Reading between the lines: linguistic cues to deception in online dating profiles. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*. ACM, 5–8.
[103] Rhiannon N Turner and Richard J Crisp. 2010. Imagining intergroup contact reduces implicit prejudice. *British Journal of Social Psychology* 49, 1 (2010), 129–142.
[104] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
[105] Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P Sheth. 2014. Cursing in english on twitter. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 415–425.
[106] Lawrence A Wenner. 1989. *Media, sports, and society*. sage.
[107] Fiona A White, Hisham M Abu-Rayya, Ana-Maria Bliuc, and Nicholas Faulkner. 2015. Emotion expression and intergroup bias reduction between Muslims and Christians: Long-term Internet contact. *Computers in Human Behavior* 53 (2015), 435–442.
[108] Thomas Wood and Ethan Porter. 2016. The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior* (2016), 1–29.
[109] Stephen C Wright, Arthur Aron, Tracy McLaughlin-Volpe, and Stacy A Ropp. 1997. The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology* 73, 1 (1997), 73.
[110] Robert Wu, Lora Appel, Dante Morra, Vivian Lo, Simon Kitto, and Sherman Quan. 2014. Short message service or disService: issues with text messaging in a complex medical environment. *International Journal of Medical Informatics* 83, 4 (2014), 278–284.
[111] InJeong Yoon. 2016. Why is it not just a joke? Analysis of Internet memes associated with racism and hidden ideology of colorblindness. *Journal of Cultural Research in Art Education* 33 (2016).
[112] Yang Yu and Xiao Wang. 2015. World Cup 2014 in the Twitter World: A big data analysis of sentiments in US sports fans' tweets. *Computers in Human Behavior* 48 (2015), 392–400.
[113] Jason Shuo Zhang, Chenhao Tan, and Qin Lv. 2018. "This is why we play": Characterizing Online Fan Communities of the NBA Teams. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 197 (2018).

Table A1. A sample of positive, negative and swear words in the Linguistic Inquiry and Word Count dictionary (LIWC [77]). Words ending with "*" match any string with the same prefix.

Positive	credit*, graced, attract*, graceful*, terrific*, bonus*, affection*, humour*, delicious*, love, openness, sweetheart*, bless*, bold*, madly, fine, friend*, hurra*, ready, trust*, secur*, won, improving, fiesta*, dynam*, toleran*, sunniest, optimal*, helpful*, neat*, enthus*, joking, favour*, giving, agreeab*, easiness, supportive*, frees*, graces, gentler
Negative	ignor*, aggravat*, unattractive, scary, attack*, offend*, grief, fright*, domina*, unfriendly, violat*, grave*, nast*, suck, shock*, sucker*, impatien*, wept, heartless*, shake*, battl*, moron*, vanity, aggress*, masochis*, unsure*, screw*, lost, losing, mocker*, envie*, sadness, nag*, timid*, afraid, hateful*, turmoil, agoniz*, obnoxious*, pain
Swear	prick*, dyke*, tit, cock, dicks, butt, bloody, dick, sob, asshole*, pussy*, screw*, suck, wanker*, mofo, fucks, shit*, bastard*, arse, butts, darn, sucked, jeez, nigger*, fucker*, arses, ass, hell, crappy, dang, motherf*, dumb*, heck, crap, tits, queer*, bitch*, sonofa*, titty, fuckin*

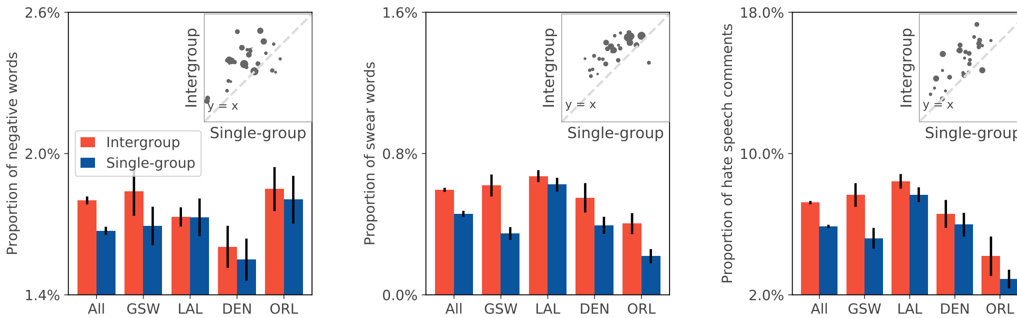
Table A2. Examples of hate speech comments detected with the automated detection model [23].

Kevin Sorbo's Hercules was such a pussy magnet
Because I want losers like you to fuck off
Same reason KAT ass rapes our team everytime we play.
Holy shit Sabonis is stuntin' like his daddy right now with these passes
That's like me saying you're a dumbass because your team is currently shit. Sorry pal, don't talk about basketball until you make the playoffs.
Melo gets NO calls ever - I've always said he should bitch more
No defense, no rebounding. Same old shit. Embarrassing
Fuck the Celtics!
now since your dumb ass sees that it doesnt make difference like ive been saying what is your excuse?
Tyler Ennis you bum. Worst player in the league.

A APPENDIX

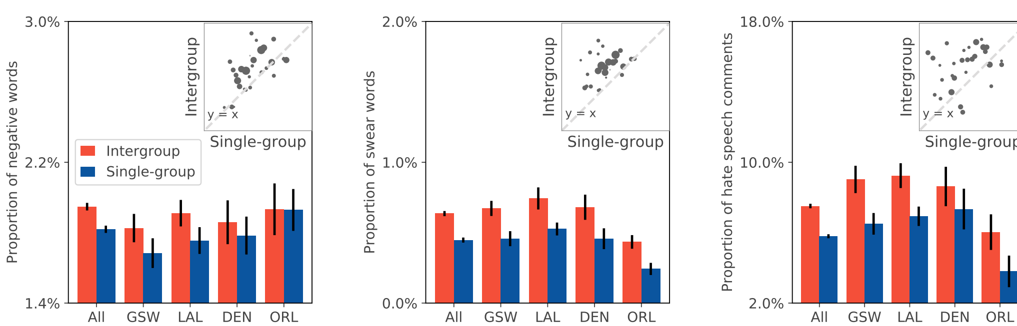
A.1 The language differences between intergroup members and single-group members are consistent when we exclude the NBA playoffs

In Section 5.1, we find that intergroup members use more negative and swear words and generate more hate speech comments compared to single-group members in the intragroup setting. It is possible that comments posted during the playoff season play a significant factor, as the activity in /r/NBA peaks in the playoff season [113]. Figure A1, A2, and A3 show that our results are consistent in all three seasons when we exclude the playoff season.



(a) Proportion of negative words.
(b) Proportion of swear words.
(c) Proportion of hate speech comments.

Fig. A1. The comparison of language usage between intergroup and single-group members in the 2018 season when excluding the NBA playoffs. Intergroup members use more negative words (two-tailed t-test, $t = 4.31$, $p < 0.001$, 95% CI=0.07% to 0.19%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 3.43$, $p < 0.001$, 95% CI=0.04% to 0.14%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 10.05$, $p < 0.001$, 95% CI=1.38% to 2.05%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$). Error bars represent standard errors.

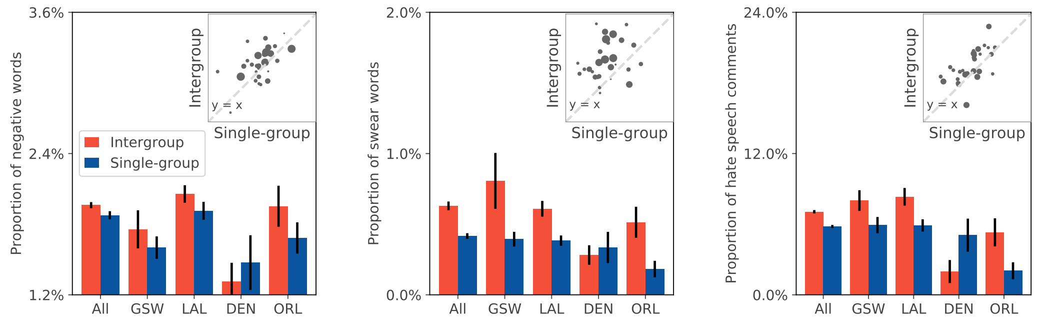


(a) Proportion of negative words.
(b) Proportion of swear words.
(c) Proportion of hate speech comments.

Fig. A2. The comparison of language usage between intergroup and single-group members in the 2017 season when excluding the NBA playoffs. Intergroup members use more negative words (two-tailed t-test, $t = 4.31$, $p < 0.001$, 95% CI=0.07% to 0.19%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 3.43$, $p = 0.002$, 95% CI=0.04% to 0.14%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 10.05$, $p < 0.001$, 95% CI=1.38% to 2.05%; 23 out of 30 teams, two-tailed binomial test $p < 0.005$). Error bars represent standard errors.

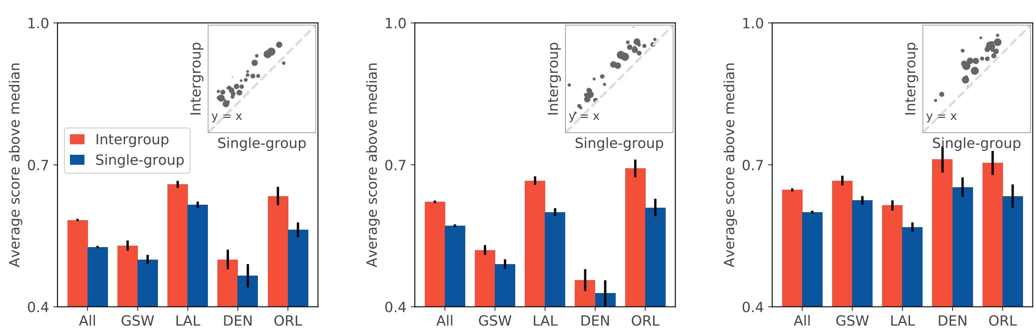
A.2 Intergroup members receive better feedback than single-group members

Figure A4 shows that intergroup members receive better feedback in the intragroup setting than single-group members in all three seasons (two-tailed t-test, $t = 15.68$, $p < 0.001$, 95% CI=0.048 to 0.062%, 29 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2018 season; two-tailed t-test, $t = 12.56$, $p < 0.001$, 95% CI=0.043 to 0.058%, 30 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2017 season; two-tailed t-test, $t = 12.43$, $p < 0.001$, 95% CI=0.046 to 0.064%,



(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

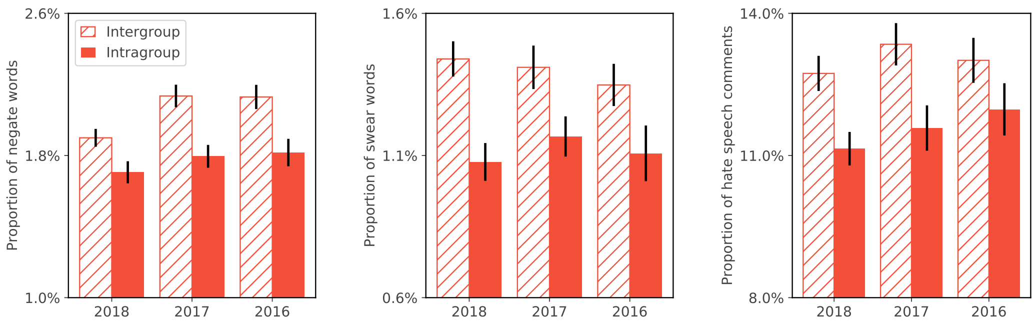
Fig. A3. The comparison of language usage between intergroup and single-group members in the 2016 season when excluding the NBA playoffs. Intergroup members use more negative words (two-tailed t-test, $t = 1.99$, $p = 0.04$, 95% CI=0.01% to 0.17%; 21 out of 30 teams, two-tailed binomial test $p = 0.04$) and swear words (two-tailed t-test, $t = 3.06$, $p = 0.002$, 95% CI=0.04% to 0.18%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 7.95$, $p < 0.001$, 95% CI=0.92% to 1.53%; 25 out of 30 teams, two-tailed binomial test $p < 0.001$). Error bars represent standard errors.



(a) The 2018 season. (b) The 2017 season. (c) The 2016 season.

Fig. A4. The comparison of feedback received from affiliated team subreddit between intergroup and single-group members in the 2018, 2017, and 2016 seasons. Intergroup members receive better feedback than single-group members in all three seasons (two-tailed t-test, $t = 15.68$, $p < 0.001$, 95% CI=0.048 to 0.062%, 29 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2018 season; two-tailed t-test, $t = 12.56$, $p < 0.001$, 95% CI=0.043 to 0.058%, 30 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2017 season; two-tailed t-test, $t = 12.43$, $p < 0.001$, 95% CI=0.046 to 0.064%, 30 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2016 season). Error bars represent standard errors.

30 out of 30 teams, two-tailed binomial test $p < 0.001$ for the 2016 season). Comment feedback is defined by whether the comment score (#upvotes- #downvotes) is above the median score of that team subreddit in that month, which accounts for the differences across subreddits. This observation suggests that using negative language is likely to draw attention in the corresponding team subreddits.

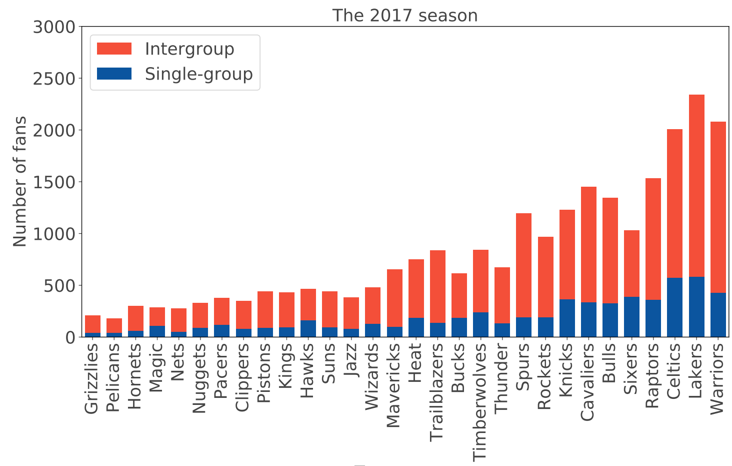


(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Fig. A5. The observation that intergroup members are more negative in the intergroup setting than in the intragroup setting is robust after controlling for topics of discussion by only considering game threads. Intergroup members use more negative words (two-tailed t-test, $t = 2.43$, $p = 0.015$, 95% CI=0.04% to 0.35% for the 2018 season; two-tailed t-test, $t = 3.76$, $p < 0.001$, 95% CI=0.16% to 0.52% for the 2017 season; two-tailed t-test, $t = 3.04$, $p = 0.002$, 95% CI=0.11% to 0.51% for the 2016 season) and swear words (two-tailed t-test, $t = 3.99$, $p < 0.001$, 95% CI=0.18% to 0.54% for the 2018 season; two-tailed t-test, $t = 2.34$, $p = 0.023$, 95% CI=0.04% to 0.32% for the 2017 season; two-tailed t-test, $t = 1.96$, $p = 0.048$, 95% CI=0.00% to 0.48% for the 2016 season) and generate more hate speech comments (two-tailed t-test, $t = 3.10$, $p = 0.002$, 95% CI=0.06% to 2.59% for the 2018 season; two-tailed t-test, $t = 2.70$, $p = 0.007$, 95% CI=0.05% to 3.05% for the 2017 season; two-tailed t-test, $t = 1.42$, $p = 0.155$, 95% CI=0.00% to 2.46% for the 2016 season) in the game threads of the intergroup setting than that of the intragroup setting. We do not compare the language usage patterns per team in this analysis, as there are teams with less than 20 members after this control. Error bars represent standard errors.

A.3 Intergroup members are more emotional in the intergroup setting than in the intragroup setting when controlling for the discussion topic

In Section 5.2, we find that intergroup members use more negative language in the intergroup setting than in the intragroup setting. This difference may occur due to the fact that more heated topics are discussed in /r/NBA than in team subreddits. To control for this factor, we further limit our comparison to the game threads in both settings. Game threads are important components of NBA-related team subreddits to facilitate game-related discussions during NBA games. In practice, each game has a game thread in the home-team subreddit, the away team subreddit, and the overall /r/NBA. Figure A5 shows the language usage difference of intergroup members in the game threads of the intergroup and intragroup setting. Only members who made comments in the game threads of both settings are included in this analysis (2118 members for the 2018 season, 1495 members for the 2017 season, and 1289 members for the 2016 season). In all three seasons, it is consistent that intergroup members use more negative words in the game threads of the intergroup setting than of the intragroup setting. We do not compare the language usage patterns per team in this analysis as there are teams with less than 20 members.



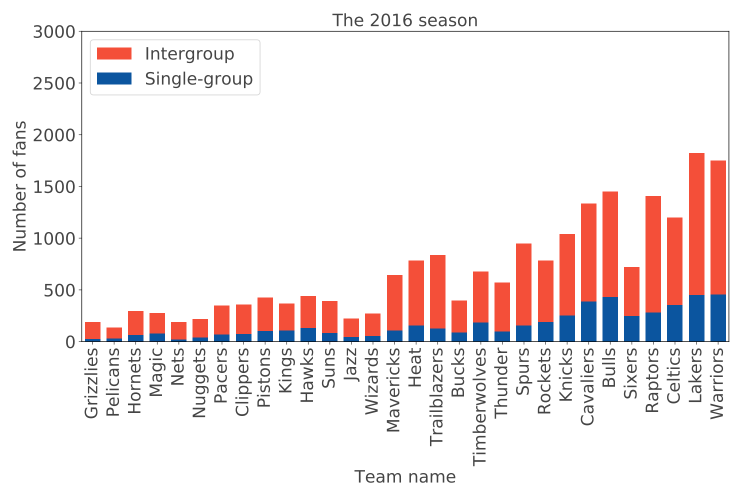
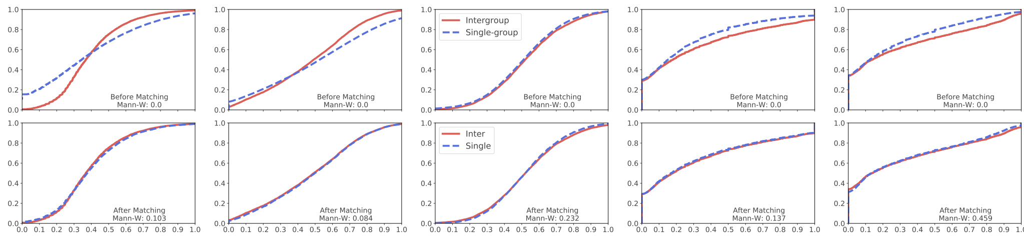
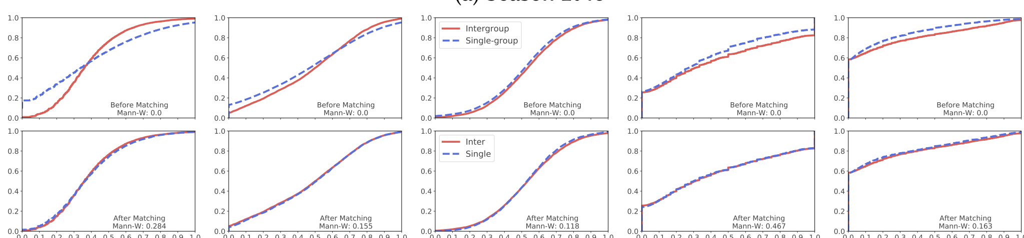


Fig. A6. The number of intergroup and single-group members affiliated with each NBA team in the 2017 and 2016 seasons. We rank 30 team subreddits by the number of subscribers each team has by the end of the 2018 season.

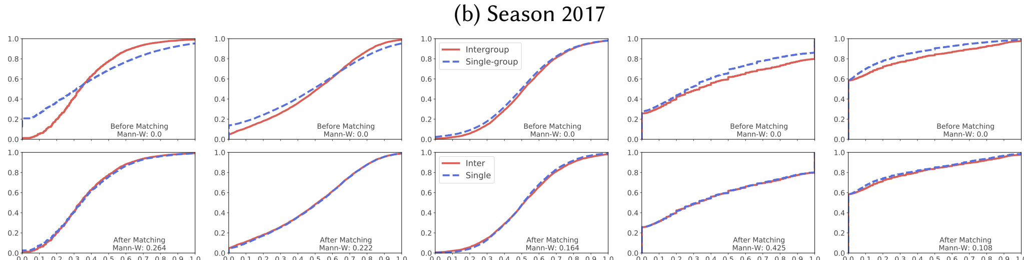
Received April 2019; revised July 2019; accepted September 2019



(a) Season 2018

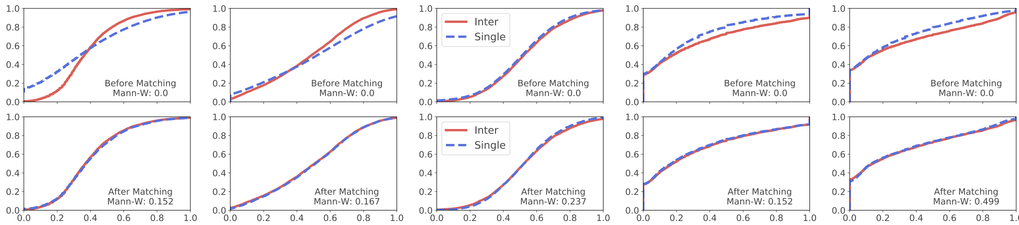


(b) Season 2017

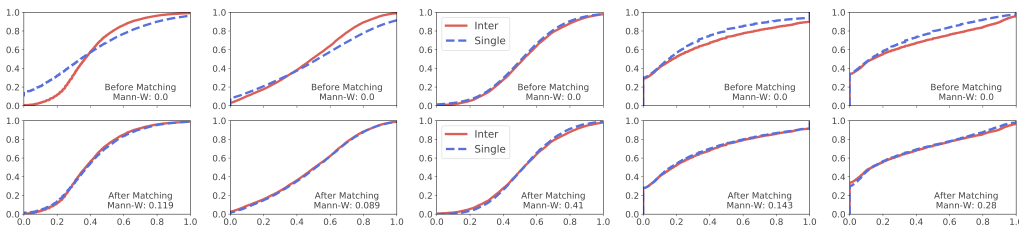


(c) Season 2016

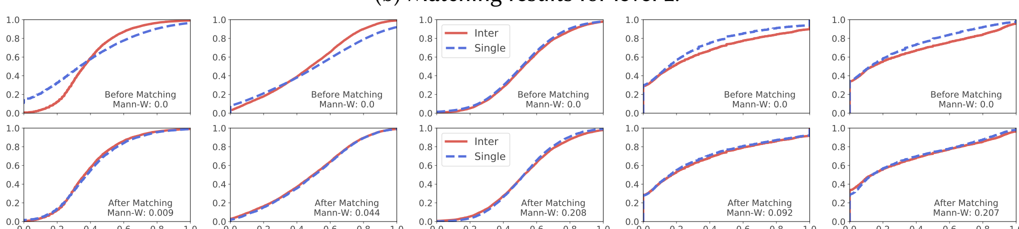
Fig. A7. Empirical cumulative distribution of each activity feature before and after the matching technique in the 2018, 2017, and 2016 seasons. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments. The corresponding p-values of the Mann-Whitney tests are also reported. Recall that a small p-value indicates that there is dependence between the treatment and control groups (relative frequencies are different). Prior to matching, each p-value is very close to 0.0. After the matching, at the 0.05 significance level (α = 0.05), we find no dependence on the group label for any activity feature observed in the matched dataset. These trends are consistent in all three seasons.



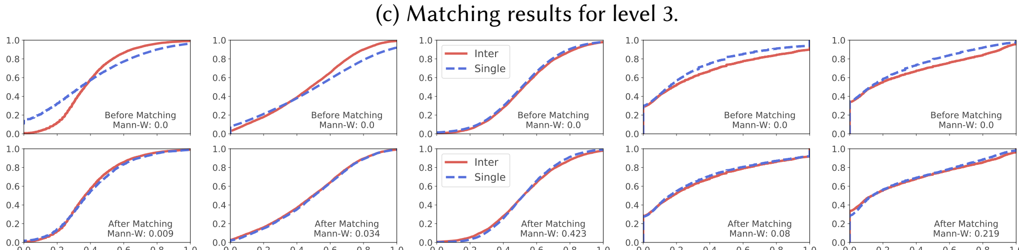
(a) Matching results for level 1.



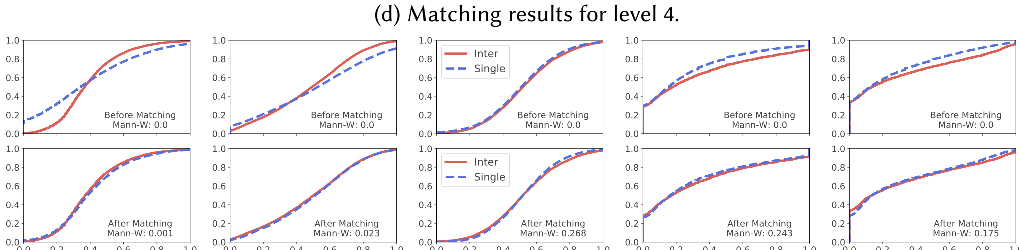
(b) Matching results for level 2.



(c) Matching results for level 3.

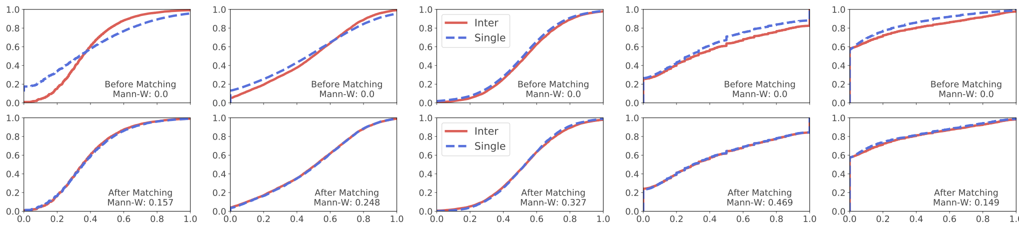


(d) Matching results for level 4.

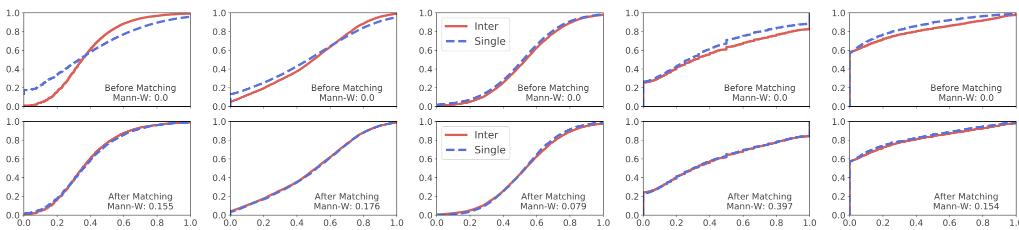


(e) Matching results for level 5.

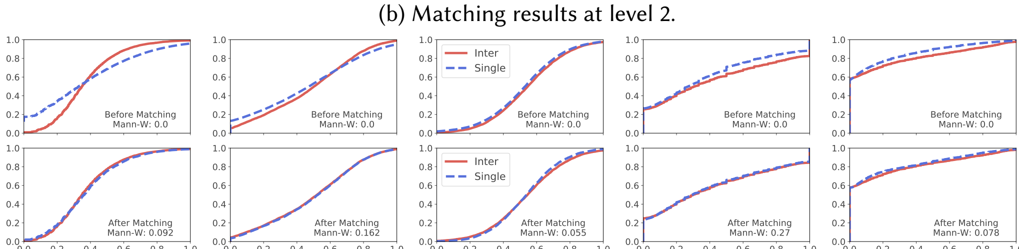
Fig. A8. Empirical cumulative distribution of each activity feature before and after the matching technique from level 1 to level 5 in the 2018 season. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments.



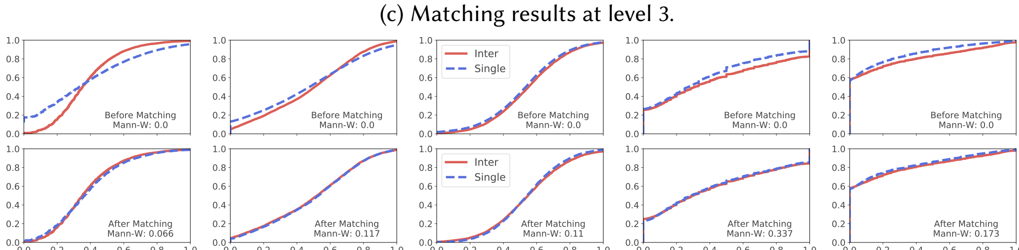
(a) Matching results at level 1.



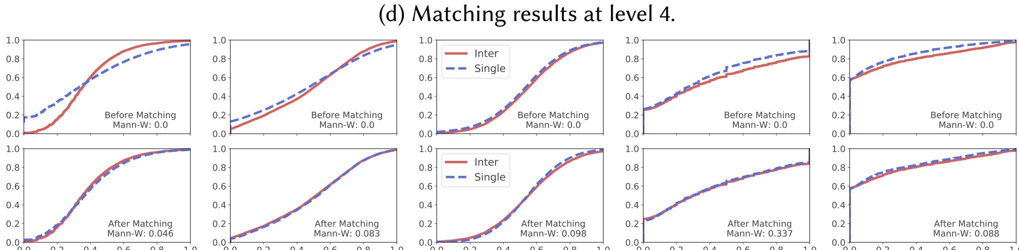
(b) Matching results at level 2.



(c) Matching results at level 3.

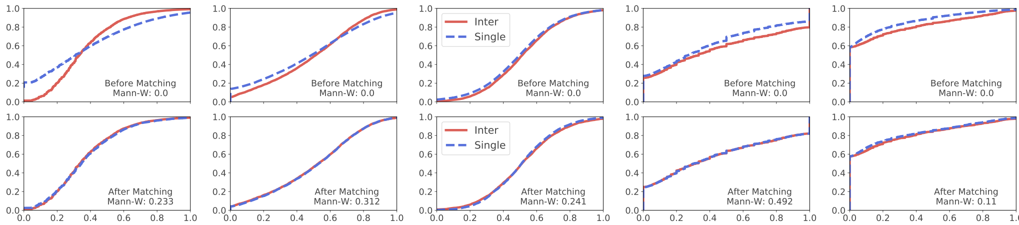


(d) Matching results at level 4.

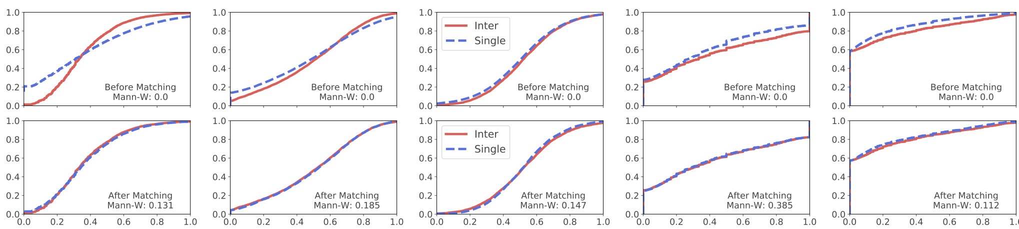


(e) Matching results at level 5.

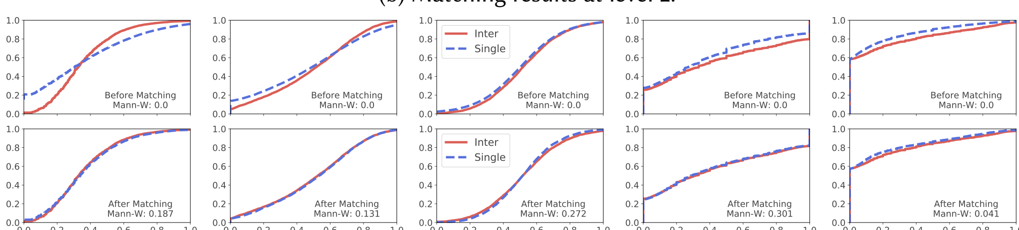
Fig. A9. Empirical cumulative distribution of each activity feature before and after the matching technique from level 1 to level 5 in the 2017 season. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments.



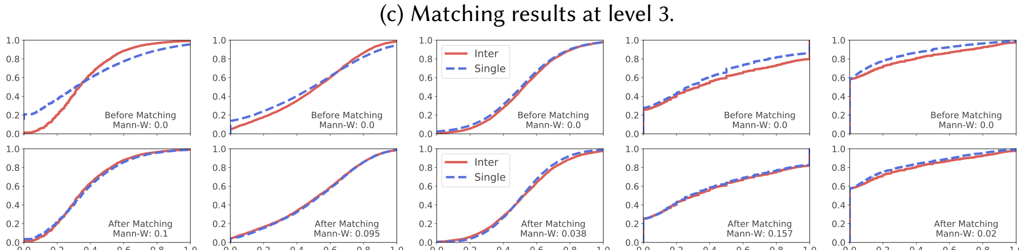
(a) Matching results at level 1.



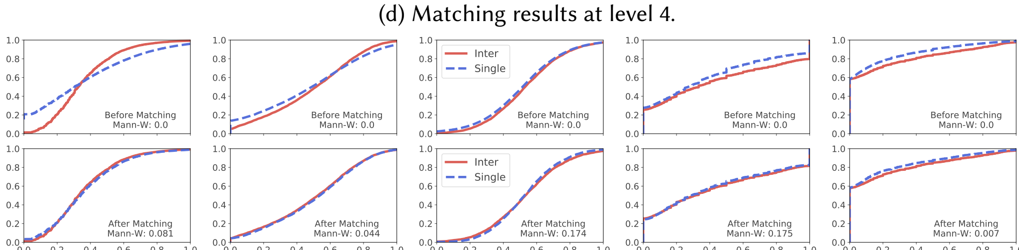
(b) Matching results at level 2.



(c) Matching results at level 3.

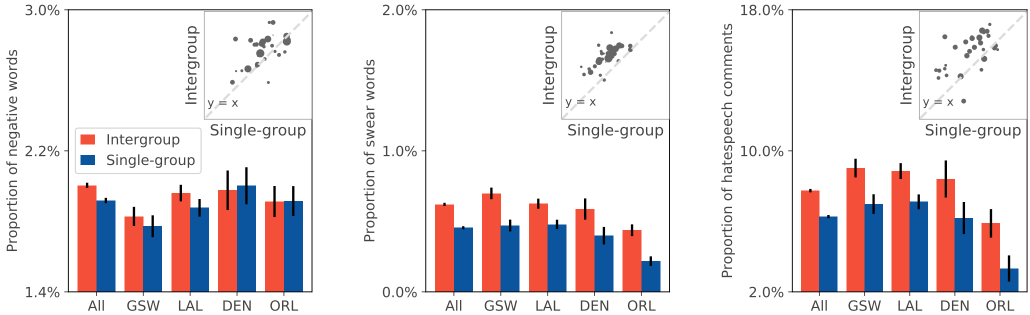


(d) Matching results at level 4.



(e) Matching results at level 5.

Fig. A10. Empirical cumulative distribution of each activity feature before and after the matching technique from level 1 to level 5 in the 2016 season. The activity features from left to right are the number of comments, the average hour gap between comments, the average comment length, the proportion of playoff comments, and the proportion of game thread comments.

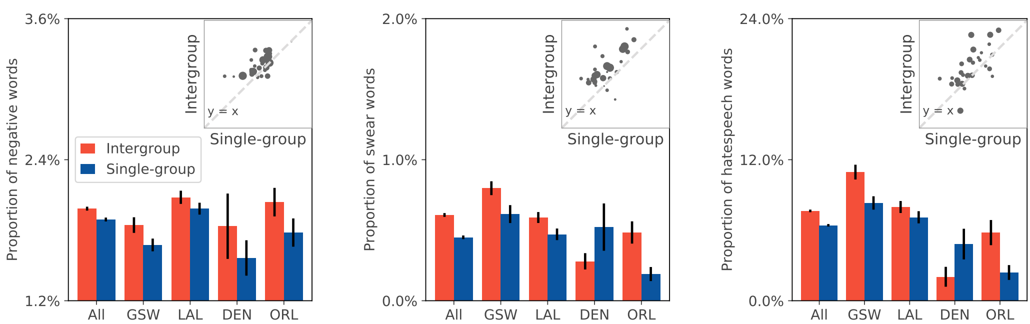


(a) Proportion of negative words.

(b) Proportion of swear words.

(c) Proportion of hate speech comments.

Fig. A11. Comparison of language usage between intergroup and single-group members in the 2017 season. Intergroup members use more negative words (two-tailed t-test, $t = 4.04$, $p < 0.001$, 95% CI=0.04% to 0.13%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$) and swear words (two-tailed t-test, $t = 4.17$, $p < 0.001$, 95% CI=0.03% to 0.10%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 11.01$, $p < 0.001$, 95% CI=1.21% to 1.74%; 24 out of 30 teams, two-tailed binomial test $p = 0.001$). Error bars represent standard errors.

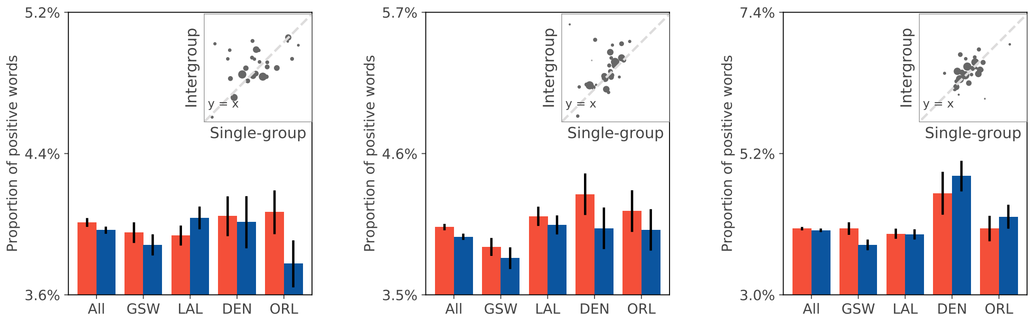


(a) Proportion of negative words.

(b) Proportion of swear words.

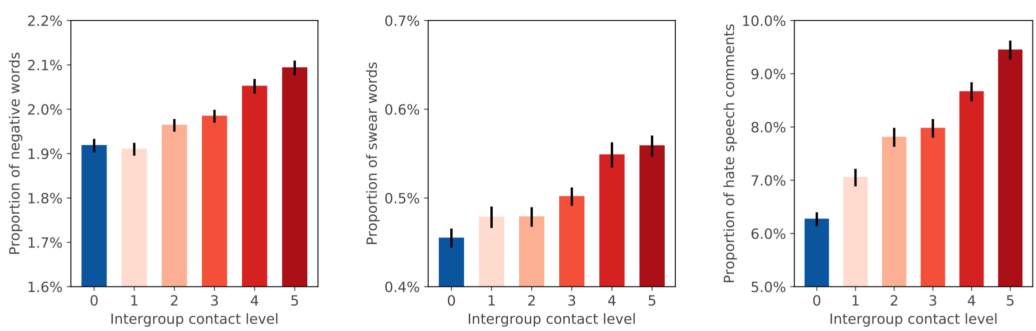
(c) Proportion of hate speech comments.

Fig. A12. Comparison of language usage between intergroup and single-group members in the 2016 season. Intergroup members use more negative words (two-tailed t-test, $t = 3.93$, $p < 0.001$, 95% CI=0.05% to 0.14%; 23 out of 30 teams, two-tailed binomial test $p = 0.005$) and swear words (two-tailed t-test, $t = 3.10$, $p = 0.002$, 95% CI=0.02% to 0.09%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 7.95$, $p < 0.001$, 95% CI=0.92% to 1.53%; 25 out of 30 teams, two-tailed binomial test $p < 0.001$). Error bars represent standard errors.



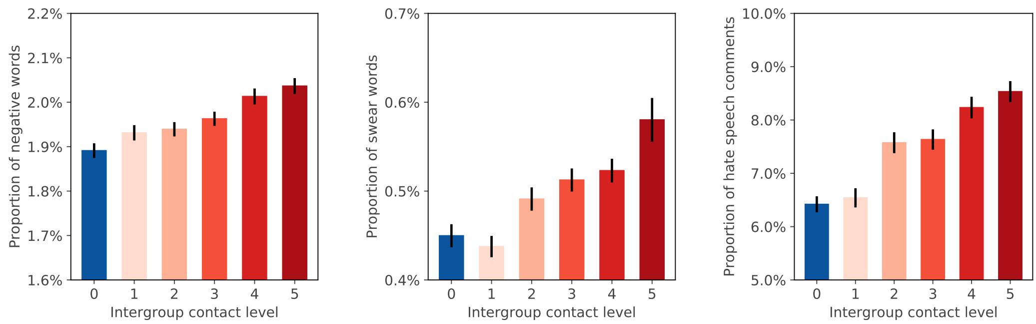
(a) The 2018 season. (b) The 2017 season. (c) The 2016 season.

Fig. A13. The comparison of positive language usage between intergroup and single-group members in the 2018, 2017, and 2016 seasons. We find no consistent trend at the 5% significance level (α = 0.05) (two-tailed t-test, $t = 1.36$, $p = 0.174$, 95% CI=0.019% to 0.107%, 16 out of 30 teams, two-tailed binomial test $p = 0.856$ for the 2018 season; two-tailed t-test, $t = 2.15$, $p = 0.03$, 95% CI=0.000% to 0.144%, 16 out of 30 teams, two-tailed binomial test $p = 0.856$ for the 2017 season; two-tailed t-test, $t = 0.66$, $p = 0.508$, 95% CI=-0.050% to 0.102%, 14 out of 30 teams, two-tailed binomial test $p = 0.856$ for the 2016 season). Error bars represent standard errors.



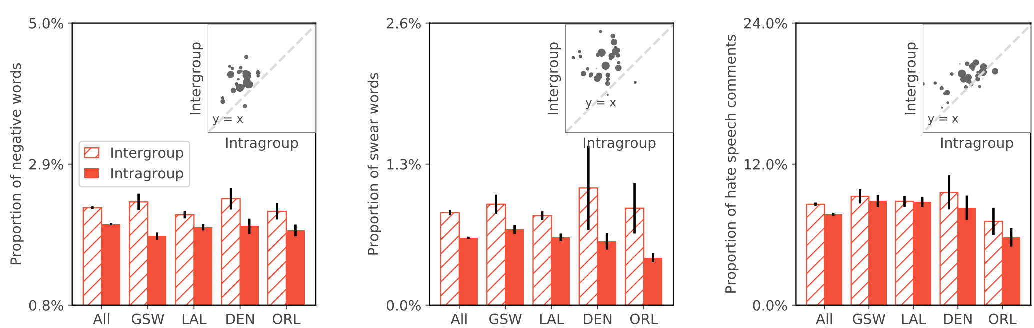
(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Fig. A14. Intragroup language usage differences of members with different intergroup contact levels in the 2017 season. x-axis represents intergroup levels determined by the fraction of comments in /r/NBA. We observe a consistent monotonic pattern in the proportion of negative words (mean = 1.92%, 1.91%, 1.96%, 1.98%, 2.05%, and 2.09%, respectively for labels from 0 to 6), swear words (mean = 0.45%, 0.48%, 0.48%, 0.50%, 0.55%, and 0.56%, respectively for labels from 0 to 6), and hate speech comments (mean = 6.27%, 7.05%, 7.81%, 8.97%, 8.67%, and 9.44%, respectively for labels from 0 to 6).



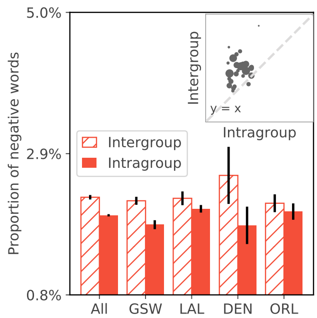
(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

Fig. A15. Intragroup language usage differences of members with different intergroup contact levels in the 2016 season. x-axis represents intergroup levels determined by the number of comments in /r/NBA. We observe a consistent monotonic pattern in the proportion of negative words (mean = 1.89%, 1.93%, 1.94%, 1.96%, 2.01%, and 2.04%, respectively for labels from 0 to 6), swear words (mean = 0.45%, 0.44%, 0.49%, 0.51%, 0.52%, and 0.58%, respectively for labels from 0 to 6), and hate speech comments (mean = 6.42%, 6.54%, 7.57%, 7.63%, 8.23%, and 8.53%, respectively for labels from 0 to 6).

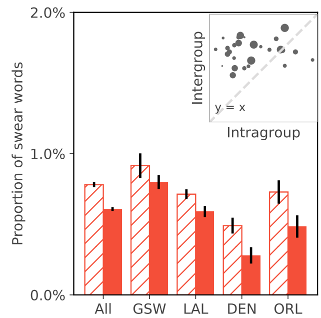


(a) Proportion of negative words. (b) Proportion of swear words. (c) Proportion of hate speech comments.

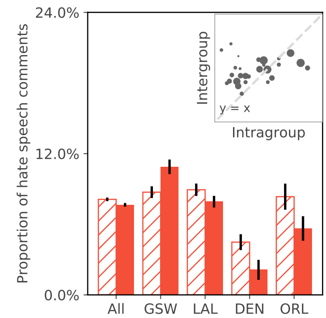
Fig. A16. Intergroup members use more negative language in the intergroup setting than in the intragroup setting in the 2017 season. They use more negative words (two-tailed t-test, $t = 9, 36$, $p < 0.001$, 95% CI=0.19% to 0.30%; 29 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (two-tailed t-test, $t = 13.80$, $p < 0.001$, 95% CI=0.29% to 0.38%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, 4.43, $p < 0.001$, 95% CI=0.48% to 1.25%; 26 out of 30 teams, two-tailed binomial test $p = 0.005$). Error bars represent standard errors.



(a) Proportion of negative words.



(b) Proportion of swear words.



(c) Proportion of hate speech comments.

Fig. A17. Intergroup members use more negative language in the intergroup setting than in the intragroup setting in the 2016 season. They use more negative words (two-tailed t-test, $t = 7.21$, $p < 0.001$, 95% CI=0.20% to 0.34%; 28 out of 30 teams, two-tailed binomial test $p < 0.001$) and swear words (two-tailed t-test, $t = 12.22$, $p < 0.001$, 95% CI=0.23% to 0.32%; 27 out of 30 teams, two-tailed binomial test $p < 0.001$) and generate more hate speech comments (two-tailed t-test, $t = 2.17$, $p = 0.030$, 95% CI=0.04% to 0.89%; 22 out of 30 teams, two-tailed binomial test $p = 0.016$). Error bars represent standard errors.