

# The Polarizing Impact of Continuous Presence on Users' Behavior

Wienke Strathern,<sup>1</sup> Angelina Mooseder,<sup>1</sup> Jürgen Pfeffer<sup>1</sup>

<sup>1</sup>School of Social Sciences and Technology, Technical University of Munich  
wienke.strathern@tum.de, angelina.mooseder@tum.de, juergen.pfeffer@tum.de

## Abstract

Online political communities offer spaces where people share and exchange political views, content, and ideas. However, people seeking political exchange online are increasingly confronted with an exclusionary intensification of discussion that no longer makes it possible to participate in constructive discourse. Online political discussion forums like r/The Donald and r/ChapoTrapHouse on Reddit have been banned recently due to the proliferation of hate speech and anti-social behavior of users. Homogeneous online discussion communities have been shown to play a key role in political polarization. Spending time in these communities tends to foster strong political positions associated with conflict. In this study we analyzed social media data from online political communities on Reddit and asked: how does presence in polarizing environments influence users' behavior? With initially equal user activity, what happens if one group of users continues to be present while the other is not? Our analysis shows that continuously present users become more active, use simpler vocabulary, and employ more abusive words in their text contributions. Our results have implications for automated moderation of polarizing online communities.

## 1 Introduction

Millions of users actively participate in online communities and social media networks offering a platform for exchanging views on any political topic. The platforms make it possible for people to network and find a community that matches their interests, convictions and political views. By offering them a space to exchange opinions, social media operators had high hopes for these platforms. However, negative online effects such as hate speech, anti-social behavior by users and offline spillover effects such as the storming of the US Capitol in January 2021, which were fomented on politically active social media forums are studied (Gallacher, Heerdink, and Hewstone 2021). The often unrestricted and anonymous environment of online discussions can become a platform for anti-social behavior, such as online abuse or harassment (Walther 2022). The drifting apart of political opinions and attitudes on platforms has been much discussed

(Kubin and von Sikorski 2021). The fact that people spend inordinate amounts of time in their close-knit social media networks—sometimes referred to as filter bubbles (Pariser 2011)—is seen as a central problem in the study of polarization (Sunstein 2009). Strong political opinions are associated with conflict, misinformation and a reluctance to engage with people and perspectives beyond one's (often narrow) own views. Given the high-speed communication of social media, its polarizing effects are even stronger.

**Research Questions** Reddit is a community driven discussion platform for political and polarizing discourse (Kane and Luo 2019; Phadke, Samory, and Mitra 2021). In a variety of political subreddits like-minded people come together to discuss current political topics. Given its openness and speed of communication with unrestricted ways of expression, discussions can be polarizing, heated, engaging, controversial, productive, creative, diverse in their facets. Due to its speed, its unrestricted flow of information and the constantly high number of contributions, users, if present, are continuously exposed to a variety of topics. Almost half of the Reddit users are heavy Internet users<sup>1</sup>, studies show that even short time effects can give first initial insights on changing behavior (Zhou et al. 2021). The question arises as to what presence in a polarizing environment does to people linguistically? Does it change the way users behave? Does it change the way users write? Regarding theories on polarization and radicalization in political settings (Grover and Mark 2019) we assume that prolonged dwelling in these polarizing environments affects users' behavior in terms of activity, thus interactions. Furthermore, if we assume that polarizing environments promote homogeneity, we expect this to be reflected in linguistic patterns of complexity. With respect to studies of online radicalization in polarizing environments, we assume that people show differences in behavior and that it is reflected in their textual contributions.

**Methods** To answer our research question methodologically, we ask: given the same level of activity, what happens if one group of users continues to be present while the other is not? In order to assess differences in behavior we measure interactions and text contributions, apply a linguistic metric

<sup>1</sup><https://www.pewresearch.org/journalism/2016/02/25/reddit-news-users-more-likely-to-be-male-young-and-digital-in-their-news-preferences/>

to determine the level of lexical diversity and apply a metric to determine the level of profane word usage in users' posts and comments.

To test our hypothesis we analyzed social media data of three famous online political communities on Reddit. We used data from three main political subreddits alongside the political spectrum to test our hypothesis: *r/The\_Donald*, a right-wing subreddit, *r/politics*, a forum for all users interested in politics, and *r/ChapoTrapHouse*, a left-wing subreddit. The explosive nature of the topic is shown by the fact that two of the subreddits we investigated, *r/ChapoTrapHouse* and *r/The\_Donald*, were banned, i.e., the discussion site was dissolved, due to hate speech, anti-social behavior, and violations against Reddits hate policy<sup>2 3</sup>. The user communities for these subreddits generated an extensive chat history of roughly 17 million posts and comments. Since these are structurally large communities, we want to know whether our hypothesis applies to both the right and the left, as well as to the moderate spectrum.

We compared the behavior of users who were continuously active to the behavior of similar users who were non-active for at least 14 days to determine 1) whether users change in activity and language and 2) whether this change is correlated with absence/non-activity. To answer our research question, we a) defined groups and b) measured posting and commenting activity for each group, the level of lexical diversity and profane word usage in their text comments. By comparing two groups and their linguistic attributes we were able to extract data-driven insights about the users' activity and linguistic changes in three online political communities. In addition, we took advantage of naturally occurring variation in the degree of overall activity of users—i.e., how frequently users posted over the time period of six months—to analyze whether behavioral changes were stronger in people who are generally more engaged in a subreddit.

**Contributions.** Based on our analysis, we derive the following conclusions:

1. Continuously present people increase their activity, show more contributions.
2. Continuously present people decrease their lexical diversity, use simpler vocabulary.
3. Continuously present people increase their profanity, use more offensive language.
4. Continuously present users who are in general more active show greater behavioral changes regarding activity and lexical diversity.

To summarize, we observe that users who are continuously present on Reddit change in activity, vocabulary and emotions, while users who are inactive for a certain time show the same level of activity, vocabulary and emotions after their absence as before.

---

<sup>2</sup>[https://en.wikipedia.org/wiki/Controversial\\_Reddit\\_communities](https://en.wikipedia.org/wiki/Controversial_Reddit_communities)

<sup>3</sup>[https://en.wikipedia.org/wiki/R/The\\_Donald](https://en.wikipedia.org/wiki/R/The_Donald)

## 2 Related Work

Our study is motivated by work discussing polarizing environments on social media (2.1) measuring activity and language (2.2) to assess psycho-linguistic behavioral attributes.

### 2.1 Polarizing Online Environments

The unrestricted way of freedom of expression in online environments can lead to hate speech and discrimination of marginalized groups, hence causing polarizing effects. A substantial body of work explores ideologically cross-cutting discussion spaces and the role of non-political spaces for political discussion. In this work the authors examine the fact that online news consumption follows a polarized pattern, observing that users' visits to news sources aligned with their own political leaning are substantially longer than their visits to other news sources (Garimella et al. 2021). Another study focused on identifying high and low consensus news posts on Twitter and presenting a method to automatically detect them (Babaei et al. 2018). In another study cross-cutting posting was tested. This determined that people are more likely to accept a news article containing conflicting views when it is delivered by a chatbot (Zarouali et al. 2021). A focus of this work was on news-link sharing. It shows that Reddit users' voting and re-sharing behaviors generally decrease the visibility of extremely biased and low factual content (Weld, Glenski, and Althoff 2021). In this paper the authors show that political conversations are less toxic in non-political subreddits (Rajadesingan, Budak, and Resnick 2021).

Several works study the effects of social media use. For example, they examine the temporary lapse from social media platforms on behavior (Allcott et al. 2020; Kovacs, Wu, and Bernstein 2021; Brown 2020) and the effects of how to make online spaces more civil (Wadden et al. 2021). Reddit has become a famous community discussion platform for political discourse. Reddit data is used to measure similarity in the commenting user bases of communities (Mamié, Horta Ribeiro, and West 2021). Another study discovers language biases encoded in the vocabulary of on-line discourse communities on Reddit (Ferrer et al. 2021). A study was conducted on analyzing online news sharing at scale to study bias a factual news on Reddit (Weld, Glenski, and Althoff 2021). This work examines the social makeup of online communities to understand the social organization of online platforms on Reddit (Waller and Anderson 2021).

### 2.2 Measuring activity and linguistic patterns

User behavior plays an important role in understanding social media platform effects. In their analysis (Jhaver et al. 2019) the authors characterized the removal explanations that are provided to Redditors and link them to measures of subsequent user behaviors. In order to better understand political engagement, this study analyzes the political interaction network on Reddit contradicting the echo chamber narrative (De Francisci Morales, Monti, and Starnini 2021). A further paper addresses the social effects of content ratings on Reddit (Davis and Graham 2021). Another study observed Redditors behavior in terms of how they interact with

### 3 Data

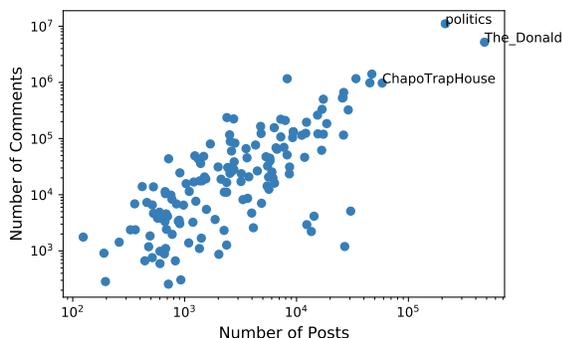


Figure 1: Correlation of number of posts and comments of 146 political subreddits

health-related messages (Silberman and Record 2021). Language, more precisely **linguistic attributes** play a key role in examining people’s behavior. Some studies have used language to record psychological stress in social media (Guntuku et al. 2019). In a work exploring the nature of political discussions in homogeneous and cross-cutting communication spaces based on interaction and linguistic patterns analysis, different behavioral patterns in homogeneous and cross-cutting communications spaces are revealed (An et al. 2019). Lexical items can be used to detect early changes in behavior (Strathern et al. 2020). In order to gain insights on how people feel more comfortable online, a study measured the effect of moderation on health in online conversations based on linguistic features (Wadden et al. 2021).

**Lexical diversity** plays a key role in measuring complexity. It reflects the complexity of vocabulary knowledge as well as the level of language proficiency. Many indices of lexical diversity have been proposed, most of which involve statistical relationships between types and tokens and which ultimately reflect the rate of word repetition. Type-Token-Ratio (TTR) provides insights into the vocabulary knowledge. This is a syntactical index that divides the number of distinct words (types) by the total number of words (tokens), computed as a running average based on consecutive 1000-word blocks of text. High values indicate texts with a heterogeneous vocabulary and linguistic structure of moderate complexity, whereas a low level of lexical diversity indicates simple terminology and little complexity (Jarvis 2013).

**Profanity** is a socially offensive use of language, its main function being to express emotions, especially anger and frustration. The purpose is to express the speaker’s emotional state and communicate that information to listeners (Jay and Janschewitz 2008). We apply a usage-based approach to study linguistic properties and functions assuming that frequency distributions in the contributions of individual users can, under certain circumstances, be interpreted regarding the writers’ underlying mental representations (Bowers and Pleydell-Pearce 2011; Schmid et al. 2021).

We describe the dynamics of three political subreddits, the dataset studied in this work, and basic preprocessing steps to filter out low-quality data.

In its structure and function, Reddit <sup>4</sup> contains elements of a discussion forum, a social network, and a news service. Users submit posts, called submissions. These can be original content, links to external content, or a combination of both. Much of the content is linked to other websites. Other users can then add comments to a post. In addition to commenting on posts, users can also rate posts and comments by upvoting them (meaning they are worthy of being seen by others) or downvoting them (meaning they should not be seen). This voting controls the display of posts and comments on the website. Overall, Reddit is divided into subreddits, which are roughly equivalent to forums or topics on other online message boards. The names of these subreddits usually describe the topic discussed (e.g., r/politics). A Pew Research Center publication gives information on the demographics of Reddit users: While only 4% of U.S. adults report using Reddit, about seven in ten of those users (78%) obtain news from the site. Overall, 2% of U.S. adults obtain news on Reddit. Both Reddit users in general and those who source news from the site tend to be young, male, and more likely to describe themselves as more liberal than the general population. About seven in ten (71%) of Reddit news users are men, 59% are between 18 and 29 years old, and 47% describe themselves as liberal, while only 13% are conservative (39% describe themselves as moderate). By comparison, of all adults in the U.S., about half (49%) are men, only 22% are between the ages of 18 and 29, and about a quarter (24%) describe themselves as liberal. Reddit news users are also heavy Internet users, with 47% reporting being online almost constantly (compared to 21% of U.S. adults overall). In this study we are interested in polarizing open online discussion forums. We would like to understand interactions and linguistic patterns. The platform offers an open space, open access and freedom of speech. There is no hierarchy – we find polarizing environments and user affiliation. Furthermore, we find informal language, a huge amount of contributions, few rules, colloquial language, and anonymity which allows for provoking behavior. Thus, the dataset is appropriate for the study of online user behavior. <sup>5</sup>

#### 3.1 Dataset description

Using the Pushift Dataset (Baumgartner et al. 2020), we collected posts and comments from 146 political subreddits from January 2018 to August 2019. Fig. 1 shows the number of posts and number of comments for these subreddits (Pearson’s  $r = 0.75$ ,  $p = 0.0$ ), with both axes being log-scaled. We choose the three subreddits with the highest posting activity for our analysis: The left-wing community r/ChapoTrapHouse, the right-wing community

<sup>4</sup><https://www.redditinc.com/>

<sup>5</sup><https://www.pewresearch.org/journalism/2016/02/25/reddit-news-users-more-likely-to-be-male-young-and-digital-in-their-news-preferences/>

r/The\_Donald, as well as r/politics, which can be described as a meeting place for politically interested users.

**Co-occurrence network.** An additional co-occurrence analysis of users in the political subreddits confirms the structural positions of these three subreddits. Fig. 2 is based on the original two-mode network data of a user being connected to a subreddit in case the user had posted in a particular subreddit. The subsequent transformation into a one-mode network of subreddits that are connected by shared users was then reduced for the visualization to show only the top five connections for every subreddit. The use of Visone’s (Baur et al. 2002) backbone visualization procedure can reveal the different political areas in Reddit’s political discussion forums. r/The\_Donald can be found in an area of the network dominated by conservative subreddits and r/ChapoTrapHouse is densely connected through shared users with a set of liberal subreddits. The central position of r/politics results from the fact that this forum is populated by users from a wide range of political views.

**Preprocessing.** For each of these subreddits, we stored all comments and posts and removed all duplicated posts, as well as posts where the author information was missing or where the contribution was created by the AutoModerator instead of a real person. While posts could include a title as well as text, comments only included text. For each post, we combined the title and text into a text field and saved the author, the day of creation and the combined text. For each comment, we saved the author, the day of creation and the text. In the following analysis we did not differentiate between comments and posts but regarded them as objects from the same type (namely contribution). For the purpose of this study, we have limited the time frame to six months and used the Reddit data from January 1 to June 30 of 2018. A brief overview of the remaining data is given in Table 1.

subreddit	Users	Comments	Posts
r/The_Donald	126,217	5,204,877	478,614
r/politics	368,056	11,081,988	213,484
r/ChapoTrapHouse	18,719	974,558	58,286

Table 1: Dataset details

## 4 Study Design

Our study is motivated by the assumption that continuously present users in polarizing environments change their behavior. Continuous interactions make the user more exposed to the existing community and less to other influences, thus being repeatedly exposed to the same opinions which reinforces existing behavior and enhances further activity. We define presence as active interactions in terms of posts and comments contributions, absence in turn is defined as no activity at all for at least 14 days. Based on these assumptions we compare the behavior of present users with the behavior of similar users with absence to determine 1) whether users change with respect to their activity and language and 2) whether this change is correlated with absence.

### 4.1 Setup

The setup of this study is depicted in Fig. 3. We have built two groups showing a similar activity level in the beginning of our study (see time frame A). While the group of present users is present during the whole analysis period, the group of absent users is absent for at least 14 days (see time frame B). We then compare how each group has changed, by comparing the behavior in time frame C to the behavior in time frame A. In the following the selection process for the present and absent group is described in more detail.

### 4.2 Defining groups

We define absence on a subreddit as inactivity in posting and commenting on a subreddit for at least 14 days (i.e., a minimum 14-day difference between the creation days of two consecutive comments by the same author). To build our *absent group* for a subreddit, we collected all cases where a user was absent from the subreddit during the six months of our data. For each user, we analyze the time frame starting with the 7 days before an absence and ending with the 7 days after the absence. This is our investigation period for the absent user. To ensure that all users were active in the 7 days before and after absence (and not absent again), we included only those users, whose first contribution in the investigation period was created less than 14 days after the preceding contribution and whose last contribution in the investigation period was created less than 14 before the following contribution. A user could have multiple entries in the absent group when being absent multiple times. For each absent user, we collected the contributions in the 7 days before absence as well as the contributions in the 7 days after absence.

We then built the set of *present users*. To ensure that the comparison between present and absent users would not be influenced by inherent activity differences in both groups, we decided to choose for each absent user a matching present user with a similar activity level. To furthermore ensure that time and external events have no effect on the comparison, we decided to analyze a present user in the same time frame as his/her matching absent user. The matching process worked as follows: for each absent user, we randomly selected a user who was present during the whole investigation period of the absent user and whose number of contributions in the 7 days before the absence of the treatment user had a minimal difference to the number of contributions of the absent user in this time frame. These users represented the present group. A user could have multiple entries in the present group, when he/she was chosen as partner for multiple treatment cases. For each present user, we collected the contributions in the 7 days before the absence of his/her matching absent user (time frame A in Fig. 3) as well as the contributions in the 7 days after the absence of his/her matching absent user (time frame C in Fig. 3). Both groups for r/The\_Donald consisted of 17,157 cases, for r/politics of 52,071 cases and for r/ChapoTrapHouse of 2,788 cases.

### 4.3 Measuring Activity

**Methods.** To analyze whether continuous interactions in a polarizing environment has an affect on posting and commenting activity, we did the following: for each user in each



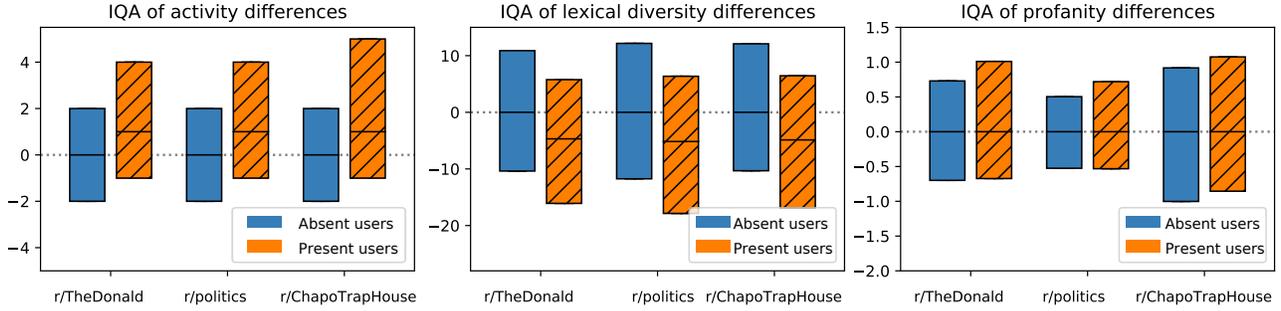


Figure 4: Interquartile range (central lines represent median) of activity, lexical diversity and profanity differences.

Differences in	Subreddit	Absent Users	Present Users	Between Groups
Activity	r\The_Donald	$Z = -2.17, p = .029$	$Z = -30.60, p < .001$ ↗	$U_{min} = 117047361, p < .001, r = .21$
	r\politics	$Z = -2.18, p = .029$	$Z = -46.20, p < .001$ ↗	$U_{min} = 1105221144, p < .001, r = .19$
	r\ChapoTrapHouse	$Z = -0.05, p = .962$	$Z = -13.27, p < .001$ ↗	$U_{min} = 3101638, p < .001, r = .20$
Lexical Diversity	r\The_Donald	$Z = -2.14, p = .033$	$Z = -32.12, p < .001$ ↘	$U_{min} = 102863734, p < .001, r = .18$
	r\politics	$Z = -2.62, p = .009$ ↗	$Z = -58.37, p < .001$ ↘	$U_{min} = 941222516, p < .001, r = .18$
	r\ChapoTrapHouse	$Z = -2.21, p = .027$	$Z = -12.58, p < .001$ ↘	$U_{min} = 2728456, p < .001, r = .19$
Profanity	r\The_Donald	$Z = -1.29, p = .197$	$Z = 16.78, p < .001$ ↗	$U_{min} = 121043421, p < .001, r = .04$
	r\politics	$Z = -0.39, p = .694$	$Z = 32.63, p < .001$ ↗	$U_{min} = 1105253057, p < .001, r = .03$
	r\ChapoTrapHouse	$Z = -0.67, p = .504$	$Z = 5.42, p = .260$	$U_{min} = 3242894, p = .009, r = .04$

Table 2: Tests for trends in differences in activity, lexical diversity or profanity before and after a period of absence/presence in each group as well as between groups. The  $Z$ -values represent the z-standardized test statistics of the Wilcoxon signed-rank tests and the  $U_{min}$ -values represent the test statistics for the Mann-Whitney U tests. Arrows indicate a statistically significant increase or decrease in activity, lexical diversity or profanity, based on the Bonferroni adjusted significance level of .017 .

ing subreddit has an affect on language use in the comments and posts made on this subreddit, we first needed to extract all users, for whom a text comparison before and after absence/presence is possible. All absent users had at least one contribution before and one after absence. All present users in our data set have made at least one contribution before the absence of their matching absent partner (thus, a user was selected, whose contribution behaviour before absence was the most similar to the absent user’s). However, present users could have zero contributions after absence, which would make a text comparison for them impossible. Therefore, we excluded all cases in the present group, where a present user had no contributions in the 7 days after absence, as well as their matching cases in the absent group. Furthermore, we excluded all cases and their matching partners if the users wrote no words in any contributions in the time frame before or after absence. This was for example the case, when users contributed only pictures or messages including only symbols and/or numbers. Both groups for r/The\_Donald included 15,850 cases, for r/politics 4,7847 cases and for r/ChapoTrapHouse 2,596 cases. As comments and posts were short in general, we combined the texts of all contributions a user has made in the 7 days before (the own or matching users) absence (time frame A in Fig. 3) to a single text. We did the same for all contributions in the 7 days after (the own or matching user’s) absence (time frame C in Fig. 3). For each text we removed special characters, punc-

tuation, empty spaces and numbers. To measure lexical diversity, we calculated for each text the percentage of unique words by all words. For each user we subtracted the diversity percentage of a text before (the own or matching user’s) absence from the diversity percentage of a text after (the own or matching user’s) absence. To measure profanity, we used a list of profane words, provided by the python package better-profanity<sup>6</sup>, and used them as a profanity dictionary. For each text we calculated the percentage of profane words by all words. For each user we subtracted the profanity percentage of a text before (the own or matching user’s) absence from the profanity percentage of a text after (the own or matching user’s) absence. We compared both groups with regards to their lexical diversity and profanity differences, using Mann-Whitney U tests. Furthermore, we assessed for each group whether their difference values would follow a symmetric distribution around zero, using Wilcoxon signed-rank test. As we conducted the test for each of the three subreddits, we used a Bonferroni adjusted significance level of  $0.05/3=0.017$ .

**Results for lexical diversity.** In the middle part of Fig. 4 we see the interquartile range of lexical diversity differences for each group, with the middle line representing the median. The statistical values of all tests can be found in Table 2. On all subreddits, continuous interaction has an effect on lexical diversity, as the difference values in the absent group were

<sup>6</sup><https://pypi.org/project/better-profanity/>

Differences in	Subreddit	Absent users	Present Users
Activity	r/The_Donald	$r = -.02, p = .014$	$r = .29, p < .001$
	r/politics	$r = -.01, p = .043$	$r = .27, p < .001$
	r/ChapoTrapHouse	$r = .01, p = .784$	$r = .27, p < .001$
Lexical Diversity	r/The_Donald	$r = .01, p = .334$	$r = -.15, p < .001$
	r/politics	$r = .01, p = .015$	$r = -.15, p < .001$
	r/ChapoTrapHouse	$r = .03, p = .207$	$r = -.12, p < .001$
Profanity	r/The_Donald	$r = -.02, p = .019$	$r = .02, p = .013$
	r/politics	$r = .00, p = .799$	$r = .02, p < .001$
	r/ChapoTrapHouse	$r = .01, p = .609$	$r = -.01, p = .489$

Table 3: Correlations between overall activity and the differences in activity, lexical diversity or profanity before and after a period of absence/presence

statistically significantly higher than the difference values in the present group. Present users (=users, who were continuously engaged in a subreddit) on all subreddits decreased in their lexical diversity over time. In contrast, absent users on r/politics significantly increased in their lexical diversity after an absence, while absent users on the other two subreddits did not change significantly.

**Results for profanity** The lower part of Fig. 4 shows the interquartile range of profanity differences for each group, with the middle line representing the median. The statistical values of all tests can be found in Table 2. On all three subreddits, continuous interaction has an effect on profanity, as the difference values in the absent group were statistically significantly lower than the difference values in the present group. Present users (=users, who were continuously engaged in a subreddit) on r/The\_Donald and r/politics significantly increased in their level of profanity. While the interquartile range also indicates an increase of profanity for present users on r/ChapoTrapHouse this was not statistically significant. Absent users in contrast, did not change their level of profanity.

**Interpretation.** Our results demonstrate that when users are present and remain constantly active, the diversity in their language decreases and their use of profanity increases.

#### 4.5 Measuring Overall Activity

In the following, we will test whether the observed effects are influenced by activity, i.e. are the effects stronger for users, who are in general more active?

**Methods.** To assess the influence of overall activity on the different patterns that present and absent users had shown, we calculated for each subreddit for each user in our groups the number of contributions he/she has made on this subreddit in the six-months observation period. The number of contributions was not normally distributed. We therefore tested for a monotonic relationship between the number of contributions and the difference values in each group using Spearman’s rank correlation coefficient, with a Bonferroni adjusted significance level of  $0.05/3=0.017$ .

**Results for activity.** In Table 3 we can see the statistical values of the correlations. On all three subreddits there is a weak positive correlation between the activity differences and the number of contributions in the present group: The more a user is generally active over the whole time pe-

riod the more he/she increases in activity if he/she remains present and continuously active. In contrast, there is no correlation between the activity differences and the number of contributions in the absent group on r/ChapoTrapHouse and on r/politics, and a significant, but negligible negative correlation between the activity differences and the number of contributions in the absent group on r/The\_Donald.

To visualize how this would affect the differences between present and absent group, we sorted all users of each group on each subreddit according to the number of contributions and built subgroups of size equal to 10% of overall group size. We then contrasted the interquartile range of present users with the interquartile range of absent users in each subgroup. The resulting image is shown in Fig. 5. The figure shows that in the group of present users on all three subreddits more active users have a higher increase in activity than less active users.

**Results for lexical diversity.** The statistical values of the correlations can be found in Table 3. On all three subreddits there is a weak negative correlation between the lexical diversity differences and the number of contributions in the present group: The more a user is generally active the more he/she decreases in lexical diversity when being continuously present in a subreddit. In contrast, there is no correlation between the lexical diversity differences and the number of contributions in the absent group on r/ChapoTrapHouse and on r/The\_Donald, and a significant, but negligible positive correlation between the activity differences and the number of contributions in the absent group on r/politics. The effect of general activity on differences between present and absent user is visualized in Fig. 6. The figure shows that in the present groups of all three subreddits more active users have a higher decrease in lexical diversity than less active users.

**Results for profanity.** The statistical values of the correlations can be found in Table 3. There is a significant, but negligible positive correlation between the profanity differences and the number of contributions in the present group on r/politics and r/The Donald, as well as no correlation on r/ChapoTrapHouse. Similarly, there is no correlation between the profanity differences and the number of contributions in the absent group on all three subreddits.

**Interpretation.** We can state that the effect of continuous presence on activity and lexical diversity is influenced by

overall activity. This means that people who are generally more active show a greater increase in activity and a greater decrease in diversity after an activity period than people who are in general less active.

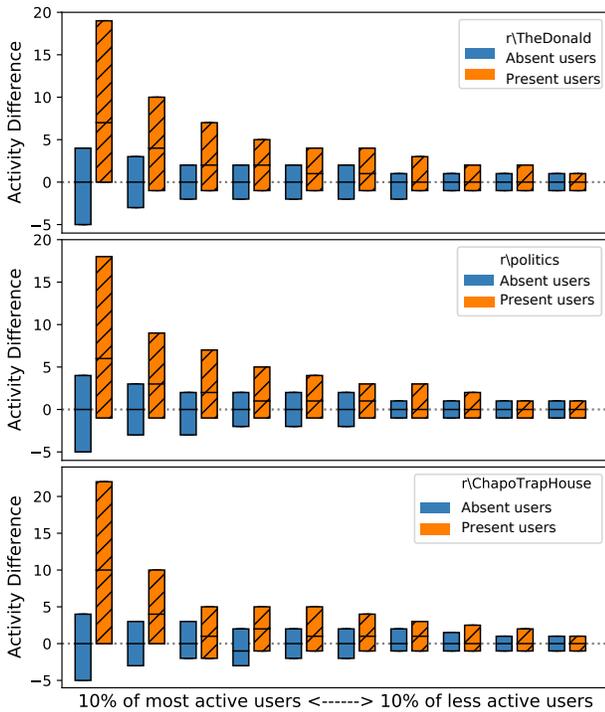


Figure 5: Activity differences of users grouped by overall activity for three subreddits.

## 5 Discussion and Conclusion.

To investigate the question of how presence in polarizing environments influences people linguistically, we analyzed social media data from three political subreddits on Reddit. We compared the behavior of continuously engaged users to the behavior of similar users who were absent for a certain time to determine 1) whether users change regarding their activity and language and 2) whether this change is influenced by absence.

For all three online political communities we found that given the same initial level of activity, users who are continuously present increase their activity and their language changes compared to users with a non-activity. We were able to determine this equally for three of the biggest online political communities on Reddit.

As a result, we can conclude that:

1. Continuously present people increase their activity, decrease their lexical diversity and increase their level of profanity.
2. The effect of continuous presence on activity and lexical diversity is stronger for people who are in general more active.

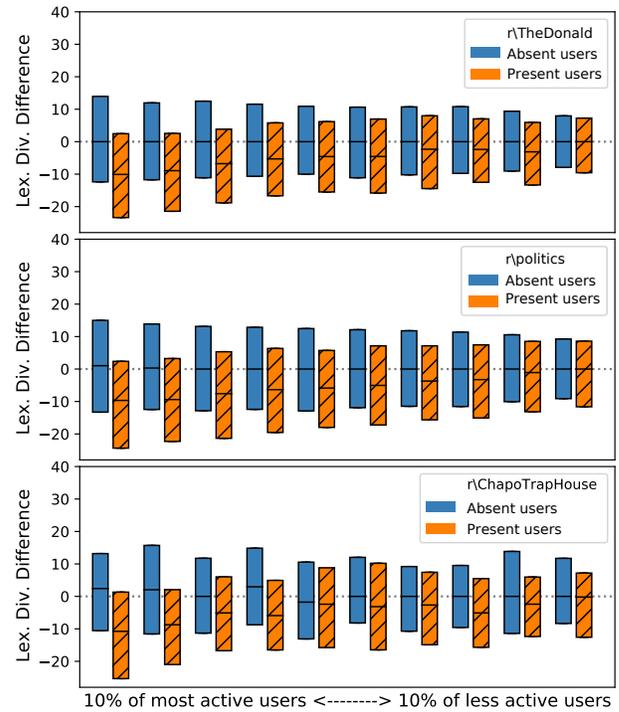


Figure 6: Lexical diversity differences of users grouped by overall activity for three subreddits.

In other words, when dwelling in polarizing environments people engage more in discussions, become simpler in terminology and more angry.

### 5.1 Limitations.

As we rely on the Pushift Dataset (Baumgartner et al. 2020), we cannot assess whether there were any missing data due to data collection errors. Using social media to study human behaviour is flawed by many challenges (Ruths and Pfeffer 2014). We are aware, however, that there were posts and comments in our dataset which were deleted by users or moderators before being retrieved by Pushift. This data was marked with a "[deleted]" in the author field and had therefore to be excluded from the analysis. From the data that were available, we leveraged word lists implemented in Python in order to gain insights into the general wording of users. Furthermore, we did not identify any obvious idiosyncrasies with regard to this platform which would prevent our findings from being generalized to other online political forums although a broader comparative study would help to determine this question. Our research represents an attempt at understanding what presence in polarizing environments does on user behavior. Follow up research could leverage contextualized word embeddings (Peters et al. 2018) to identify which mentions of potentially harmful words demand attention. In addition, future work could complement our linguistically-driven approach by comparing lexical choices (Wang and Culotta 2019).

## 5.2 Implications and Future Work.

In our analysis we compared the behavior of present users to the behavior of similar users with absence to determine whether users' behavior changes in polarizing environments. We found that users participate more, become simpler in vocabulary and use more offensive language. Spending inordinate amounts of time in close-knit environments is seen as a central problem and our results confirm slight changes. Based on this we would like to enrich mass media research on user activity and on short-term intervention techniques. Media consumption leads to changes in behavior and being active online in communities is one form of media consumption. News that is shared and active social interaction in form of discussions influence consumption behavior and this in turn influences the selection of news and interaction. That said, reinforcement effects in behavior and attitudes can occur through selective exposure (Berelson, Gaudet, and Lazarsfeld 1944; Katz 2001).

The question we want to pose here is how the mass media stream of information consumption can be interrupted and whether non-consumption in turn can influence behavior. Hence, we open the discussion for possible (automatic) content moderation techniques to counteract reinforcement effects. In particular, we think of time outs as in sports or the *dead cat* in British parliament discussions<sup>7</sup>. However, little is known about the effectiveness of concrete interventions such as blocking single users temporarily in online political communities. The core motivation and function of short-term interventions could be to slow down activity of single group participants to balance emotions and avoid escalation. Can a time-out have a de-escalating effect? Future work could include experiments which actively induce a user's absence to measure the effect of time-outs caused by others. Furthermore, it would be interesting also to examine the long-term effects of online political conversations. When do these discussions lead to constructive discourses and conversations not resulting in a ban from these platforms and, given that conversations often turn on a particular instance in the discourse, what are the micro-level linguistic traits that enhance long-term stability?

## References

Allcott, H.; Braghieri, L.; Eichmeyer, S.; and Gentzkow, M. 2020. The Welfare Effects of Social Media. *American Economic Review* 110(3): 629–676.

An, J.; Kwak, H.; Posegga, O.; and Jungherr, A. 2019. Political discussions in homogeneous and cross-cutting communication spaces. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 68–79.

Babaei, M.; Kulshrestha, J.; Chakraborty, A.; Benevenuto, F.; Gummadi, K. P.; and Weller, A. 2018. Purple Feed: Identifying High Consensus News Posts on Social Media. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 10–16.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The Pushshift Reddit Dataset. In *Pro-*

*ceedings of the International AAAI Conference on Web and Social Media*, 830–839.

Baur, M.; Benkert, M.; Brandes, U.; Cornelsen, S.; Gaertler, M.; Köpf, B.; Lerner, J.; and Wagner, D. 2002. Visone : Software for Visual Social Network Analysis. In Mutzel, P., ed., *Graph drawing : 9th international symposium*, 463–464. Springer.

Berelson, B.; Gaudet, H.; and Lazarsfeld, P. F. 1944. *The People's Choice. How the voter makes up his mind in a presidential campaign*. Duell, Sloan & Pearce.

Bowers, J. S.; and Pleydell-Pearce, C. W. 2011. Swearing, Euphemisms, and Linguistic Relativity. *PLoS ONE* 6(7): e22341.

Brown, A. J. 2020. “Should I Stay or Should I Leave?”: Exploring (Dis)continued Facebook Use After the Cambridge Analytica Scandal. *Social Media + Society* 6(1).

Davis, J. L.; and Graham, T. 2021. Emotional consequences and attention rewards: the social effects of ratings on Reddit. *Information, Communication & Society* 24(5): 649–666.

De Francisci Morales, G.; Monti, C.; and Starnini, M. 2021. No echo in the chambers of political interactions on Reddit. *Scientific Reports* 11(1): 2818.

Ferrer, X.; Nuenen, T. v.; Such, J. M.; and Criado, N. 2021. Discovering and Categorising Language Biases in Reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 140–151.

Gallacher, J. D.; Heerdink, M. W.; and Hewstone, M. 2021. Online Engagement Between Opposing Political Protest Groups via Social Media is Linked to Physical Violence of Offline Encounters. *Social Media + Society* 7(1).

Garimella, K.; Smith, T.; Weiss, R.; and West, R. 2021. Political Polarization in Online News Consumption. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 152–162.

Grover, T.; and Mark, G. 2019. Detecting Potential Warning Behaviors of Ideological Radicalization in an Alt-Right Subreddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 193–204.

Guntuku, S. C.; Buffone, A.; Jaidka, K.; Eichstaedt, J. C.; and Ungar, L. H. 2019. Understanding and Measuring Psychological Stress Using Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 214–225.

Jarvis, S. 2013. Capturing the Diversity in Lexical Diversity. *Language Learning* 63: 87–106.

Jay, T.; and Janschewitz, K. 2008. The pragmatics of swearing. *Journal of Politeness Research* 4,2.

Jhaver, S.; Birman, I.; Gilbert, E.; and Bruckman, A. 2019. Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator. *ACM Transactions on Computer-Human Interaction* 26(5): 31:1–31:35.

Kane, B.; and Luo, J. 2019. Do the Communities We Choose Shape our Political Beliefs? A Study of the Politicization

<sup>7</sup>[https://en.wikipedia.org/wiki/Dead\\_cat\\_strategy](https://en.wikipedia.org/wiki/Dead_cat_strategy)

- of Topics in Online Social Groups. In *IEEE International Conference on Big Data*, 3665–3671.
- Katz, E. 2001. Lazarsfeld’s Map of Media Effects. *International Journal of Public Opinion Research* 13(3): 270–279.
- Kovacs, G.; Wu, Z.; and Bernstein, M. S. 2021. Not Now, Ask Later: Users Weaken Their Behavior Change Regimen Over Time, But Expect To Re-Strengthen It Imminently. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14.
- Kubin, E.; and von Sikorski, C. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 45(3): 188–206.
- Mamié, R.; Horta Ribeiro, M.; and West, R. 2021. Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube. In *13th ACM Web Science Conference 2021*, 139–147. Association for Computing Machinery.
- Pariser, E. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin Press.
- Peters, M.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; and Zettlemoyer, L. 2018. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, 2227–2237.
- Phadke, S.; Samory, M.; and Mitra, T. 2021. What Makes People Join Conspiracy Communities? Role of Social Factors in Conspiracy Engagement. *Proc. ACM Hum.-Comput. Interact.* 4(CSCW3).
- Rajadesingan, A.; Budak, C.; and Resnick, P. 2021. Political Discussion is Abundant in Non-political Subreddits (and Less Toxic). In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 525–536.
- Ruths, D.; and Pfeffer, J. 2014. Social Media for Large Studies of Behavior. *Science* 346(6213): 1063–1064.
- Schmid, H.-J.; Würschinger, Q.; Fischer, S.; and Küchenhoff, H. 2021. That’s Cool. Computational Sociolinguistic Methods for Investigating Individual Lexico-grammatical Variation. *Frontiers in Artificial Intelligence* 3.
- Silberman, W. R.; and Record, R. A. 2021. We Post It, U Reddit: Exploring the Potential of Reddit for Health Interventions Targeting College Populations. *Journal of Health Communication* 26(6): 381–390.
- Strathern, W.; Schoenfeld, M.; Ghawi, R.; and Pfeffer, J. 2020. Against the Others! Detecting Moral Outrage in Social Media Networks. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 322–326.
- Sunstein, C. R. 2009. *Going to extremes: How like minds unite and divide*. Oxford University Press.
- Wadden, D.; August, T.; Li, Q.; and Althoff, T. 2021. The Effect of Moderation on Online Mental Health Conversations. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 751–763.
- Waller, I.; and Anderson, A. 2021. Quantifying social organization and political polarization in online platforms. *Nature* 600(7888): 264–268.
- Walther, J. B. 2022. Social media and online hate. *Current Opinion in Psychology* 45.
- Wang, Z.; and Culotta, A. 2019. When Do Words Matter? Understanding the Impact of Lexical Choice on Audience Perception Using Individual Treatment Effect Estimation. *AAAI* 33(1): 7233–7240.
- Weld, G.; Glenski, M.; and Althoff, T. 2021. Political Bias and Factualness in News Sharing across more than 100,000 Online Communities. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 796–807.
- Zarouali, B.; Makhortykh, M.; Bastian, M.; and Araujo, T. 2021. Overcoming polarization with chatbot news? Investigating the impact of news content containing opposing views on agreement and credibility. *European Journal of Communication* 36(1): 53–68.
- Zhou, X.; Rau, P.-L. P.; Yang, C.-L.; and Zhou, X. 2021. Cognitive Behavioral Therapy-Based Short-Term Abstinence Intervention for Problematic Social Media Use: Improved Well-Being and Underlying Mechanisms. *Psychiatric Quarterly* 92(2): 761–779.