frontiers in Political Science

ORIGINAL RESEARCH
published: 02 November 2021
doi: 10.3389/fpos.2021.741605

Over-Time Trends in Incivility on Social Media: Evidence From Political, Non-Political, and Mixed Sub-Reddits Over Eleven Years

Qiusi Sun[1]*, *Magdalena Wojcieszak*[1] *and Sam Davidson*[2]

[1]Department of Communication, University of California, Davis, CA, United States, [2]Department of Linguistics, University of California, Davis, CA, United States

Incivility in social media has become a major concern of the public, who perceive uncivil online interactions to be both widespread and increasing. This study provides a descriptive account of incivility dynamics over the past 11 years by examining the trends of incivility in three main categories of social media interactions: political, mixed, and non-political. Using longitudinal data from Reddit that accounts for 95% of the entire Reddit universe across 11 years and relying on the combination of supervised machine learning models and traditional statistical inference, the study found that incivility consistently represents around 10% of total Reddit comments. Additionally, political groups tend to be more uncivil, and discussions in mixed groups that are not overtly political but nevertheless discuss politics are less uncivil than in political groups. We also found that the fluctuations of incivility correspond to offline events and platform-specific policies.

Keywords: incivility, social media, machine learning, reddit, political discourse, online discussion

OPEN ACCESS

Edited by:
Alessandro Nai,
University of Amsterdam, Netherlands

Reviewed by:
Julia Partheymueller,
University of Vienna, Austria
Lukas F. Stoetzer,
Humboldt University of Berlin, Germany

***Correspondence:**
Qiusi Sun
qssun@ucdavis.edu

Specialty section:
This article was submitted to Elections and Representation, a section of the journal Frontiers in Political Science

Received: *15 July 2021*
Accepted: *15 October 2021*
Published: *02 November 2021*

Citation:
Sun Q, Wojcieszak M and Davidson S (2021) Over-Time Trends in Incivility on Social Media: Evidence From Political, Non-Political, and Mixed Sub-Reddits Over Eleven Years. Front. Polit. Sci. 3:741605. doi: 10.3389/fpos.2021.741605

INTRODUCTION

Scholars and observers worry that public debates in the United States are growing increasingly uncivil. Politicians attack their opponents, partisans report unprecedented hostility toward opposition-party supporters (see Iyengar et al., 2019), and partisan media describe the opposing party as Nazis or Communists (Berry and Sobieraj, 2013) and feature "in your face" debates (Mutz and Reeves, 2005). Concerns with incivility often pertain to the Internet. Incivility is pervasive in online communities (Reader, 2012). In 2018, 84% of Americans reported having experienced incivility online, and those who did encountered it roughly 11 times a week (KRC Research, 2018). These encounters can have negative effects. For example, the use of and exposure to incivility generates anger, anxiety, or mental distress, and can lead to aggression (Gervais, 2015) and hostile communication (Groshek and Cutino, 2016). In addition, incivility can drive users away from online discussions and lead to general dissatisfaction with public discourse (Anderson, et al., 2014; Bauman, et al., 2013; Moor, et al., 2010; Ransbotham, et al., 2016).

In this project, we aim to address a fundamental descriptive question regarding over-time variations in incivility across a range of online communities. We rely on the most comprehensive longitudinal dataset of Reddit comments from 2008 to 2019.[1] and a combination of computational

[1]We note that we have 13 years of Reddit data (i.e., 2006–2019), yet only posts and comments starting in 2008 can be analyzed for our purposes. This is because there were only 2 sub-reddits in 2006 and 4 in 2007, all created by administrators. Given that the first political sub-reddit (politics) was created in August 2007, the comparison among sub-reddit categories was done from 2008.

methods (i.e., a neural, BERT-based classifier to capture incivility in an incredibly large corpus of data, see Davidson et al., 2020) and traditional statistical inference (e.g., ANOVA and student t-test) to provide a descriptive account of online incivility 1) over-time, 2) across different contexts of online discussions (i.e., political, mixed, and non-political), and 3) as influenced by external events.

Our extensive data show that the volume of incivility increased with the overall increase in the volume of online exchanges, but its proportion remained rather constant across the years, oscillating at roughly 10%. Consistent with the general observations, discussions about politics generate consistently more incivility than non-political and mixed discussions. That said, when aggregated across the years, incivility in gaming communities that sometimes discuss politics is significantly higher than in other groups, even explicitly political ones. Supporting worries about the difficulty of cross-party exchanges, politically heterogeneous online communities–where liberals and conservatives meet–generate more incivility than politically homogeneous liberal or conservative communities. Moreover, fluctuations in incivility are affected by platform-level policies and external events.

INCIVILITY ON SOCIAL MEDIA

There is some conceptual and operational ambiguity in existing literature on incivility and related concepts under the umbrella of toxic, offensive, or intolerant speech (see Kim et al., 2020 and Rossini, 2020 for recent reviews). Sometimes incivility is used to refer to impoliteness or negativity. Yet, unlike impoliteness, incivility is seen as "individual behaviors that threaten a collective founded on democratic norms" (Papacharissi, 2004, p. 271). And unlike negativity, which can be delivered in both civil and uncivil ways and target an issue or an individual (Brooks & Geer, 2007), incivility actively demonstrates a lack of respect (Gervais, 2017) and is said to be detrimental to deliberative debate and reduce deliberative potential of offline or online conversations (Gervais, 2017). Recent work differentiates between uncivil and intolerant speech, with the former including discourse that goes against accepted social norms and the latter being discourse that promotes discrimination, derogation, and violence (Rossini, 2020).

Here, we do not address the distinction between incivility and other related concepts, nor do we test its democratic effects. We follow Coe et al. (2014), seeing incivility as "features of discussion that convey disrespectful tone toward the discussion forum, its participants, or its topics" (Coe, et al., 2014, p. 660). Accordingly, we adapt the operational definition of incivility as speech that includes "name-calling, mean-spirited or disparaging words directed at a person or a group of people, an idea, plan, policy, or behavior; using vulgarity, profanity, improper language and pejorative remarks about the way a person communicates" (Coe et al., 2014, p 660). As such, our project includes and empirically captures speech that merely counters social norms, e.g., name-calling, as well as the arguably more problematic intolerant speech that can be hateful toward social groups.

Macro Trends

Many observers lament declines in the quality of public discourse in the United States (Anderson et al., 2014; Santana, 2014) and some scholars are concerned that the affordances of social media platforms, such as anonymous or pseudonymous communication, have led to increases in incivility and its normalization in the online public sphere (Leurs and Zimmer, 2017; Theocharis et al., 2020). And yet, systematic evidence of these potential increases in incivility on social media platforms is still limited. Research on the temporal dynamics of incivility mostly focuses on Twitter - a platform used by a minority of American adults (22%; Pew Research Center, 2021) - and typically during certain contentious times and/or salient political events. The resulting evidence is mixed. For instance, Siegel et al. (2018) find no constant increases in incivility on Twitter during the 2016 presidential election and its aftermath; rather, their data suggest random spikes in incivility unrelated to external events. In contrast, analyzing longitudinal data from Twitter after the 2016 presidential election, Theocharis et al. (2020) show that the prevalence of uncivil tweets mentioning Members of the US Congress is rather stable and spikes in incivility correspond to political events (e.g., a white nationalist rally) and policy debates (e.g., healthcare). Yet in other work looking at Reddit (2021) find a sharp increase in incivility in political sub-reddits during the 2016 presidential campaign. Although those studies provide important insights into the dynamics of incivility on social media platforms, the timeframes analyzed are rather short and it is not clear whether extant worries regarding growing incivility and its normalization are warranted. By examining a much longer time span of nearly the universe of online expressions on Reddit, one of the most popular social media platforms, this project offers a macro level panorama of variations in online expressions of incivility. We first ask: RQ1: Has there been an increase in incivility, in the aggregate, on Reddit between 2008 and 2019?

Contextual Influences

In addition to offering systematic evidence on whether, and the extent to which, incivility increased on social media over the past 11 years, our major contribution lies in testing these variations across different kinds of groups. Different topics and community cultures in online groups, which are developed by niche interests and user engagement, may promote, or discourage uncivil behavior (Massanari, 2017). With different discourse dynamics, it is possible that the variations in incivility differ across various types of discussions, political and non-political alike. Our project is, to our knowledge, the first to differentiate expressions of incivility in political versus non-political groups, and, furthermore, across various categories of political groups (e.g., liberal, conservative, or heterogeneous), groups focusing on non-political issues (e.g., fashion, gaming), and also groups where users touch on both (e.g., discussing global warming in sub-reddits dedicated to cars; see Wojcieszak and Mutz, 2009).

Political, Mixed, and Non-political Discussions

Extant concerns with, and past work on, incivility mostly focuses on political incivility. This work finds substantial amounts of incivility in the comment sections of news websites (i.e., around 20% of comments were found to be uncivil in online newspaper comment sections, Coe et al., 2014) and on social media platforms (e.g., around 9% in political comments on Reddit, Nithyanand et al., 2017b, between 15 and 20% on Twitter et al., 2020). Yet, the focus on incivility in political spaces is rather narrow given that many Americans see politics as complex, boring, or overly divisive (Greenlee, 2014; Jacoby, 2018) and avoid information about news and politics altogether (Feldman et al., 2013; Guess, 2021; Prior, 2007; Wojcieszak et al., in press). Accordingly, most users do not discuss politics online (Barberá et al., 2019) and do not follow any political accounts on social media (Thorson and Wells, 2016; Eady et al., 2019). Clearly, examining strictly political incivility or incivility in overtly political spaces misses a large part of the online information and communication ecosystem.

For one, the nature of online discussion is never clear-cut, and people do engage in political exchanges in groups organized around non-political topics (Wojcieszak and Mutz, 2009). There, users connect with others based on shared non-political interests (e.g., following the same celebrity or being parents) and yet encounter politics inadvertently (e.g., when a celebrity endorses a politician on their Facebook page or a parenting sub-reddit discusses funding for education) (Wojcieszak and Mutz, 2009; Fletcher and Nielsen, 2018; Silver and Andrey, 2019). We refer to these groups as *mixed*, those where politics is *not* the central purpose but where users nevertheless engage in political talk. Even though users report encountering disagreement when political discussions emerge in non-political spaces (Wojcieszak and Mutz, 2009), research found these mixed groups generated less incivility than explicitly political discussions (Rajadesingan et al., 2021). After all, once people establish a shared interest, they may be more open to potential disagreements when politics emerges and engage with others more politely and with an open mind.

Second, as aforementioned, most people do not go online to exchange political information and may also shy away from discussions that entail any political topics altogether. Accordingly, the most popular online groups on social media platforms pertain to entertainment. For instance, the most followed Facebook pages are Facebook App, Samsung, and Cristino Ronaldo, focusing on topics such as games, technology, and celebrities (Wikipedia contributors, 2021a). Similarly, among the top ten most followed Twitter accounts, eight are celebrities, and only one (Barack Obama) is a political figure (Wikipedia contributors, 2021b). The same pattern is found in YouTube and Reddit, with all top 10 most subscribed YouTube channels and eight sub-reddits being entertainment (Baer, 2021; Wikipedia contributors, 2021c). Given the popularity of non-political spaces, we attend to these largely overlooked discussions. Even though there may be important topical differences between non-political groups, as we detail below, on average these groups may not entail as much name-calling, personal attacks, or disparaging or mean-spirited language as the political or even the mixed communities. One could expect the members of groups focused on movies, celebrities, pets, or technology to be bonded by common fandom (Seregina and Schouten, 2017) and *a priori* more favorable toward one another due to shared interests.

In sum, although mixed and non-political discussions may be less uncivil than political ones, this idea remains untested. Similarly, it is not clear whether fluctuations in incivility would differ across political, mixed, and non-political groups. If, as some fear, online discourse is increasingly uncivil, we would see growth in incivility across these three types of groups. If, however, the shared interests and common ground matter to online discourse, the trends would be less pronounced in mixed and especially in non-political groups. Given the lack of clear-cut directional expectations and the largely descriptive nature of our work, we ask: *R*Q2: Have there been changes in incivility between political, non-political, and mixed groups?

Specific Types of Online Discussions

Ideologically homogeneous vs heterogeneous political and mixed groups

To portray the tested dynamics comprehensively, we offer a nuanced differentiation within political and mixed as well as non-political groups. First, we distinguish between political and mixed groups that are ideologically homogeneous versus heterogeneous. Considering the current polarized climate in the US, discussions between people who hold different views may be substantially more uncivil than discussions between people with similar political affiliations (in that Democrats may clash with Republicans and liberals may call conservatives names). That said, ideologically homogeneous groups could also entail high levels of incivility (in that Democrats/liberals could unite against former President Trump or Republicans/conservatives could bash the policies of President Biden, for instance). Research suggests that ideologically homogeneous networks may cultivate beliefs in conspiracy theories or foster extremist attitudes (Warner and Neville-Shepard, 2014); these beliefs and attitudes may result in strong and emotional opinion expression, which, in turn, could lead to incivility (Stevens, 2021). In short, whether ideologically homogeneous or heterogeneous political and mixed discussions are more uncivil is not only unexamined but also unclear. Our next question, therefore, asks: RQ3: Have there been changes in incivility between ideologically homogeneous and heterogeneous groups?

Conservative vs Liberal homogeneous groups

Within ideologically homogeneous political and mixed groups, we attend to expressions of incivility in liberal and conservative groups. Previous studies on group identity and norms show that conservatives and liberals follow different social norms for incivility (Rains et al., 2017) and see incivility differently; for instance, conservatives are less likely to perceive messages as uncivil (Kenski et al., 2020). In addition, Donald Trump's presidency may have encouraged or normalized incivility among conservatives (e.g., during Trump's election, there was more incivility in conservative sub-reddits; Nithyanand et al., 2017a). Thus, conservatives may be more likely to express

incivility as they may see it as a usual or more accepted way of expression than liberals. On the other hand, several studies showed that on social media platforms liberals were more likely to "like." or "thumb-up" uncivil comments (Rains et al., 2017; Kim et al., 2020), indicating liberals agree with or endorse uncivil expressions; this may lead liberals to express uncivilly to gain agreement from their peers. And yet, a study on unacceptable and uncivil behavior in US politics finds that Republicans and Democrats react in similar ways to uncivil messages (Muddiman, 2021). We therefore ask: RQ4: Have there been changes in incivility between politically liberal and conservative groups?

Different non-political topics

Lastly, we examine whether incivility levels differ across various topics within the mixed and non-political groups, testing discussions revolving around entertainment, sports, lifestyle, and technology, among others (as detailed below). Some of these topics may touch on individual identity, in a way similar to political stance (e.g., sports or gaming, Vale and Fernandes, 2018; Murphy, 2004) and thus generate heated discussions that may lead to uncivil discourse. Inasmuch as, say, fans of the Dallas Cowboys see the Philadelphia Eagles as a rival, discussions about sports could be more uncivil than those about politics. Furthermore, certain hobby communities have "geek." cultures where incivility may be a norm (Massanari, 2017). For instance, participants in gaming communities may bash others for losing or call them names for poor performance (Shen et al., 2020). In short, non-political groups discussing distinct topics might differ in the volume and fluctuations in incivility. Also, some of those communities (e.g., sports, gaming) may be similar to explicitly political groups. These questions remain unaddressed in extant work. RQ5: Have there been changes in incivility between different non-political groups?

External Events

In testing these questions, we attend to the extent to which external events may influence the prevalence of and changes in incivility in the online public sphere. The aforecited research on political incivility suggests that controversial issues and events may lead ordinary citizens to express their opinions, lead to emotional engagement, and trigger uncivil expression (Theocharis et al., 2020). That is, fluctuations in incivility on social media may be triggered by offline events. Yet, because extant work mostly focuses on elections and/or specific short time periods, we do not know whether other events could lead to spikes in uncivil interactions online during non-election years and across different categories of online groups. Also, the implementation of various regulatory policies by social media platforms could be seen as an external event that influences users' behavior (Buntain et al., 2021). For instance, an analysis of YouTube's implementation of a policy regarding conspiracy-oriented channels showed a sharp and consistent change in trends of harmful content. Such policies serve to classify and regulate inappropriate behaviors and content and may lead to an increase or decrease in incivility (Blackwell et al., 2017). We thus investigate the relation between online incivility and offline events, both socio-political and also platform specific. RQ6: Have any specific external events triggered increases in incivility?

METHODS

Reddit

We rely on online behavioral data from Reddit, a social media platform with over 330 million users globally (Alexa, 2019) and 222 million in the US alone (Lin, 2021). Reddit is the only social media platform (apart from YouTube) that saw statistically significant growth since 2019 (Pew Research Centre, 2021) and a steady growth in its user base since its inception. For example, from 2013 to 2019, the annual growth rate of monthly active users ranged from 21.42% (2014) to 47.06% (2017, Curry, 2021). Reddit is the ninth most visited website globally (Top, 2018) and the tenth most popular site in the US. Clearly, users' expressions therein are important to study.

As in other social media platforms, Reddit allows users to post content and discuss various issues in individual communities, which it calls "sub-reddits." A sub-reddit is a specific community dedicated to a particular topic where users can post a link, create a post, or comment on others' posts. Each sub-reddit has its own unique rules, moderators, and themes for submissions. Currently there are more than 2.8 million sub-reddits, and more than 130,000 are active (receiving at least five comments a day, Lin, 2021). Those sub-reddits are of three privacy levels: public, restricted, and private. Any user can join and post in a public sub-reddit, but they can only join but not post in a restricted sub-reddit until the moderator approves. Private sub-reddits usually have rules governing admittance; users receive an invitation once they meet the admission requirements.

Several features of the platform are relevant to our focus on incivility. For one, unlike Facebook or Twitter, Reddit's core aspect is anonymity. Based on its privacy policy and its support for individual freedom of expression (Reddit, 2021), Reddit protects users' identity and does not require real-name or identity verification. Although this could result in an uninhibited trolling, toxicity, or hate speech on the platform, Reddit has several mechanisms in place to prevent this from happening. Most sub-reddits have community guidelines developed by the creator and also moderators that explicitly forbid incivility, toxicity, trolling, personal attacks, or other problematic language in posts and comments. For instance, r/MachineLearning emphasizes "Be nice, no offensive behavior, insults, and attacks." as its first rule, and r/AskReddit also requires users to "be respectful to other users at all times and conduct your behavior in a civil manner." The community rules are reinforced by both automatic tools called automods and human moderators. As a proactive tool, automods can remove and report posts and comments with inappropriate external links, words, and phrases. In addition, sub-reddit members are encouraged to report and downvote problematic posts and comments. Both auto and human reports go directly to sub-reddit moderators, who can remove the posts and comments that go against the sub-reddit's rules and guidelines. In addition,

administrators can remove content, ban users or even close down an entire sub-reddit based on their regular review of content and user reports.

Prior to 2015, Reddit had no specific anti-harassment policy, taking actions such as banning a user or taking down a sub-reddit only when certain concerns became public and received media attention (e.g., closing down of r/beatingwoman for violence against women and sharing users' private information or r/TheFappening for posting hacked celebrity pictures); it announced its anti-harassment policy in May 2015. Reddit defined any behavior that makes users feel unsafe and shut users out of the conversation as uncivil (e.g., menacing someone and directing abuse at a user or a group). The then-developed user reporting system allowed human moderators and administrators to decide whether a comment and a user should be removed or prohibited (before that, users could only report content or groups by contacting administrators).

Furthermore, in 2019, Reddit invited bystanders (e.g., regular users not involved in the reported issues) to provide a third-person point view on harassment reports. In addition, Reddit introduced machine learning tools to help organize and identify more severe cases. In 2020, in response to the George Floyd Protests, the policy was strengthened and further enforced. So far, Reddit still mostly relies on human judgement to identify any communities, users, or comments that go against its anti-harassment policy. Reddit's hands-off administration on the one hand and its gradually strengthened anti-harassment policy on the other hand make it a perfect platform to observe the natural flow of uncivil interactions.

We accessed all Reddit content from the beginning of Reddit.com (December 2005) up to December 2019 on PushShift's Reddit data using Google BigQuery.[2]. In total, this yielded over 6.68 billion comments. Annually, the number of unique users commenting ranged from 23,793 (in 2006) to 80,788,041 (in 2019) (M = 19,401,466, SD = 26,733,025), and the number of comments ranged from 417,184 (in 2006) to 1,663,587,081 (in 2019, M = 477,154,362, SD = 526,611,725). In order to offer comprehensive evidence on the over-time fluctuations of incivility on Reddit, we identified the most popular sub-reddits, which represented 95% of the total Reddit comments each year. We did that by 1) the number of comments in the sub-reddit and 2) the number of users who posted in the sub-reddit. This has resulted in 9,355 sub-reddits that were most popular across the years. We therefore account for 95% of the entire Reddit universe. Among all identified sub-reddits, yearly comments in a sub-reddit ranged from 1,215 (in 2006) to 84,457,656 (in 2019, M = 202,786, SD = 1,173,944), and yearly unique users ranged from 78 (in 2006) to 12,424,518 (in 2019, M = 33,162.3, SD = 161,665.3).

Sub-Reddit Annotation

Our core questions pertain to the differences in incivility between political, non-political, and mixed sub-reddits, and also ideologically homogeneous (liberal or conservative) and heterogeneous (liberal and conservative) political and mixed sub-reddits. We eliminated non-English and banned sub-reddits and also those English-speaking sub-reddits that were specifically non-US (e.g., sub-reddits from or discussing Australia, Canada, India, or the UK).[3] resulting in 8,458 sub-reddits for analysis (90.41% of all identified sub-reddits). We developed a coding manual to categorize each sub-reddit accordingly, as detailed below. Sub-reddits that discussed politics and news explicitly (e.g., r/politics, r/news) were categorized as political, while those revolving around non-political issues (e.g., r/nba, r/gaming) were categorized as non-political. In addition, the mixed category included sub-reddits whose purpose is not to discuss politics but where people discuss political issues (e.g., r/AskReddit, r/pics). In addition to these three categories, we classified the political and mixed sub-reddits into politically homogeneous or heterogeneous sub-reddits, and the former into liberal or conservative sub-reddits. Politically homogeneous sub-reddits were those where the majority of posts and comments were in favor of liberal/left/Democratic or conservative/right/Republican ideas, figures, and policies (e.g., r/BlueMidterm 2018; r/Conservative, r/proguns). In turn, heterogeneous sub-reddits were those where posts and comments had mixed perspectives (e.g., some comments supporting and other comments opposing the Democratic/Republican Party, or posts expressing both sides of an issue, such as r/news or r/PurplePillDebate). The politically homogeneous sub-reddits were further categorized as liberal (i.e., those supporting Democratic/liberal ideology and/or discussing socio-political issues from the Democratic/liberal perspective) or conservative (i.e., those supporting Republican/conservative ideology and/or discussing socio-political issues from the Republican/conservative perspective).

To address our question regarding incivility in non-political spaces, we also identified ten types of non-political sub-reddits based both on their overt purpose and content. Sub-reddits about games (video games, board games, etc.) and gaming services were categorized as Games; Entertainment category contained all sub-reddits about movies, TV programs, celebrities, and other entertainment; sub-reddits about sports, teams and athletes were categorized as Sports; Health sub-reddits included all that discussed physical and mental health; Music category included sub-reddits discussing music, instruments, and musicians; Technology sub-reddits were those discussing science and technology developments and education; sub-reddits about pets and animals were categorized as Pets/Animals; Lifestyle/Fashion category contained sub-reddits about beauty, food, clothing, design, models, and lifestyle; and all sub-reddits dedicated to creating and sharing memes were categorized as Memes. The remaining sub-reddits were categorized as Others.

[2]https://pushshift.io/using-bigquery-with-reddit-data/

[3]As an additional exploratory analysis, we describe the aggregate over-time trends in incivility for the non-US, English speaking sub-reddits in **Supplementary Appendix SE**. We find the over-time trend and the proportion of incivility in all main categories was similar to those in the US sub-reddits.

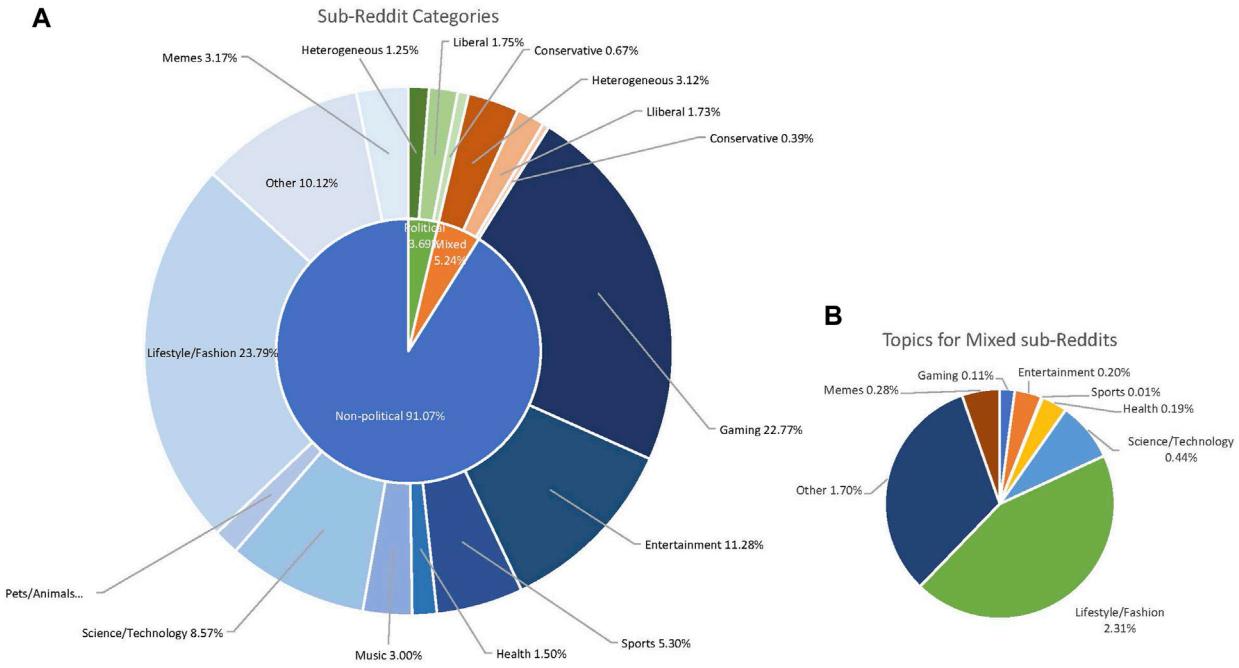

FIGURE 1 | Sub-reddit Categories with Percentage. Note. **(A)** The inner circle represents the distribution of main categories: political, mixed, and non-political with percentage of total number of sub-reddits. The outer circle represents the second level categories for each main category with percentage of total sub-reddits. The second level categories include heterogeneous, liberal, and conservative for political sub-reddits; heterogeneous, liberal, and conservative for mixed sub-reddits; gaming, entertainment, sports, health, science/technology, lifestyle/fashion, memes, and others. **(B)** The circle represents the topics in mixed sub-reddits with the percentages of total number of sub-reddits, including gaming, entertainment, sports, health, science/technology, lifestyle/fashion, memes, and others. In addition, out of political sub-reddits, heterogeneous takes 33.97%, liberal is 47.76%, and conservative is 18.27%. For mixed sub-reddits, 59.59% are heterogeneous, 32.96% are liberal, and 7.45% are conservative.

Seven trained coders labeled 8,458 sub-reddits (see **Supplementary Appendix SA** for our detailed coding procedure and inter-coder reliability). **Figure 1** shows the distribution of the categories and **Supplementary Appendix SB** presents specific examples. We identified 312 political sub-reddits (3.69% of total sub-reddits), of which 66.03% (206) were politically homogeneous and 33.97% (106) heterogeneous. Among the homogeneous sub-reddits, 72.33% 149) were liberal and 27.67% 57) conservative. Further, 443 sub-reddits were classified as mixed (non-political with at least 40% of posts and/or comments pertaining to politics; these comprised 5.24% of total sub-reddits). Among the mixed sub-reddits, 40.41% 179) were ideologically homogeneous (146 liberal, 33 conservative) and 59.59% 264) heterogeneous. The remaining 7,703 sub-reddits were non-political (91.07% of total sub-reddits), with the largest groups of non-political sub-reddits being lifestyle and fashion (n = 2012, percentage = 23.79%), followed by games (n = 1926, percentage = 22.77%) and entertainment (n = 954, percentage = 11.28%).

Incivility Annotation and Classifier

To classify Reddit content as uncivil or not, we developed and validated an incivility classifier. A coding manual was developed based on previous research (Coe et al., 2014), and three new trained coders labelled Reddit comments with binary labels as civil or uncivil. Uncivil comments were those that included 1) name-calling, mean-spirited or disparaging words directed at a person, or a group of people; 2) aspersion, mean-spirited or disparaging words directed at an idea, plan, policy, or behavior; 3) vulgarity, profanity or language that would not be considered proper; 4) pejorative or disparaging remark about the way in which a person communicates. For instance, comments such as "It's OK, you'll hit puberty one day." or "you're a dumbass for simplifying the issue and trying to jump right into the helm of the 'y'r all hypocrites' bandwagon." were coded as uncivil. Our approach accounted for both the content as well as the targets of incivility to create a comprehensive dataset for model building. Coders received five runs of coding exercises, with overall inter-coder reliability resulting in a Fleiss's kappa of 0.663, and then moved on to individual coding. A final set of 4,000 stratified sampled comments from each year was randomly assigned to coders, and the individual coding and training coding were together used for supervised model building.

In order to automatically identify incivility, we decomposed the task into three steps. We first developed neural binary classifiers built on top of large transformer-based language models, namely BERT (Devlin et al., 2018). First, a pretrained BERT model was further pretrained for domain adaptation on 3 million unlabeled Reddit comments using a masked language modeling objective. Then the model was fine-tuned for four epochs on 5,000 human-labeled comments with 10% of the data set aside for training validation and 1,000 coded

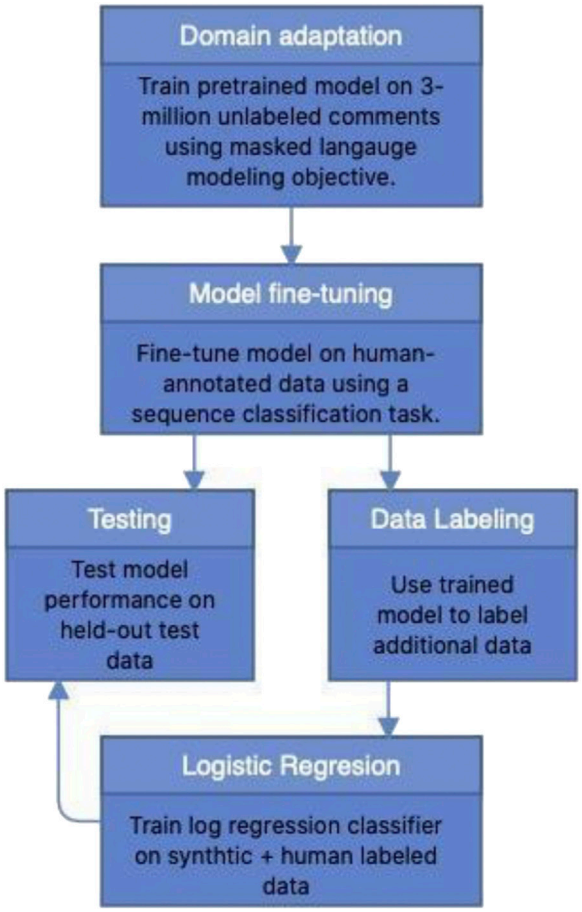

FIGURE 2 | Computational Framework Flow Diagram. Note. The graph depicts the steps of the machine learning process. Both BERT and DistilBERT followed the first two steps, but only DistilBERT was used for data labeling. The generated labeled data from the DistilBERT model was used for logistic regression.

comments set aside for model testing (see Davidson et al., 2020). The final F1-score.[4] for the classification model was 0.786. Next, we tried to improve computational performance by utilizing DistilBERT (Sanh et al., 2019), a more compact version of BERT trained using a model distillation technique. The final F1-score of the DistilBERT model was 0.802. Considering the large scale of our dataset, using BERT or DistilBERT models to classify more than 10 years' Reddit data would be both time-consuming and computationally and financially expensive. To address this constraint, in the third step, a logistic regression classification model was trained using 5 million Reddit comments labeled by our fine-tuned DistilBERT model, in addition to the smaller human-annotated dataset. The final logistic regression model achieves an F1-score at 0.779 which is similar to the performance of our BERT and DistilBERT models, and our model error falls within the 95% CI of [0.0297, 0.0547]. **Figure 2** gives an overview of the computational framework. and **Supplementary Appendix SC** offers details of model building procedures. **Supplementary Appendix SD** presents examples from both human annotation and machine classification.

RESULTS

Macro Trends

The overall yearly trends in the prevalence of incivility relative to the total content contributed is shown as the red line in **Figure 3**. Between 2008 and 2019, total incivility - depicted with the red line - fluctuated between 8 and 12%, an estimate that is largely consistent with evidence from Twitter (Theochairs, et al., 2020; Siegel et al., 2018). After slight decreases in the general proportion of incivility until about 2015, when the total proportion of comments classified as uncivil reached the lowest point of 8.84%, incivility has been gradually increasing since, with its levels rising to around 10% in 2016 and 2017. We note, however, that this increase was not dramatic and that the proportion of comments categorized as uncivil did *not* return to the high, pre-2015 levels of around 12%, which is when Reddit initiated its anti-harassment policy and banned several sub-reddits promoting incivility and hateful speech. Addressing RQ1, we note that the proportion of incivility fluctuates only slightly, with a current upward trend, and can be affected by the policies of social media platforms.

Political, Mixed, and Non-political Discussions

Are there variations in incivility across political, mixed, and non-political groups? Addressing RQ2, political groups–the yellow line in **Figure 3**—contain the highest proportion of incivility among the three major categories (i.e., political, non-political, mixed) across all the years analyzed, with the percentage oscillating between 10 and 17%. Results from one-way ANOVA ($F = 32.095$, $p < 0.001$) showed a significant difference among these categories, and post hoc Tukey's HSD indicates that incivility in political groups is significantly higher than in mixed (diff = 0.025, $p < 0.001$) and non-political groups (diff = 0.040, $p < 0.001$). It is in the political groups that we observe the steepest increase in incivility after 2015, likely due to the highly contentious 2016 presidential elections.[5] Incivility in political groups increased by 33.12% between 2015 and 2017 (see also Nithyanand et al., 2017a) and has been growing gradually since 2017.

In *mixed* groups, where participants discuss political and non-political issues, the proportion of incivility ranged between 11 and

[4]F1-score is a measurement of model accuracy for binary classification, which is calculated from precision and recall. Precision is the number of true positives (the incidents which are 1 and also identified as 1) divided by the number of all positives, while recall is the number of true positives divided by the sum of true positives and false negatives (the incidents which are 1 but identified as 0 by the machine). F1-score is the harmonic mean of precision and recall.

[5]Incivility also peaked in 2009. Based on the examination of a random sample of uncivil political comments in 2009, this increase may be due to discussions of equality and medical care.

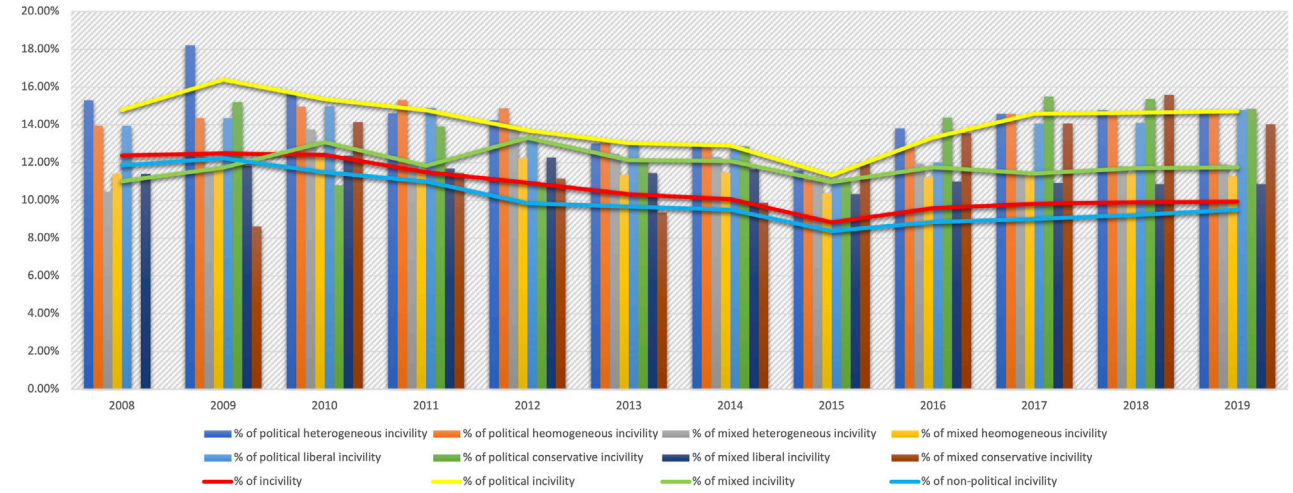

FIGURE 3 | Yearly Incivility Proportion from 2008 to 2019. Note. The percentage shown in the graph is the proportion of incivility content over relative total content.

13%, which is significantly higher than in non-political groups (t = -2.940, df = 12, p < 0.01) and significantly lower than in political groups (t = 5.333, df = 12, p < 0.001). But at some points, such as during the 2015–2016 period, incivility in mixed groups spiked, reaching levels of incivility similar to that in political groups. The temporal variations in incivility in *non-political* groups are similar to the total trends in incivility and those in political sub-reddits. As could be expected, the proportion of incivility to overall content in these groups is significantly lower than overall proportion across all the sub-reddits (t = 8.045, df = 12, p < 0.001), as well as that in political and mixed groups.

To shed more light on the variations between 2015 and 2019, we also analyzed monthly data. **Figure 4A** depicts large variations in incivility, yet the relative proportion of incivility in general and also in mixed and non-political groups remains stable. Consistent with the yearly trend, the incivility proportion in political groups (F = 2048.534, p < 0.001) is significantly higher than in mixed (diff = 0.065, p < 0.001) and non-political groups (diff = 0.064, p < 0.001). Notably, however, political incivility increased with several spikes. The peak in July 2016 can be linked to the 2016 Democratic National Convention and the early email leak of the Democratic National Committee (which can be confirmed by the boost of incivility in the liberal group, See **Figure 4C**). The observed spike in May 2017 overlapped with several actions related to the investigation of Russian interference in the 2016 US election, including the Great America Committee and dismissal of James Comey. The last peak in June 2018 corresponds to several protests against the family separation policy.

Ideological Homogeneous vs Heterogeneous Political and Mixed Discussions

RQ3 asked about incivility in ideologically homogeneous vs heterogeneous groups. **Figure 3**, which also summarizes the trends for political and mixed categories of ideologically homogeneous/heterogeneous groups from 2008 to 2019 in bars, shows that incivility in all four categories oscillated between 10 to 20%, reaching its lowest levels in 2015, and gradually increasing since then. Although incivility in ideologically heterogeneous political groups, where users encounter others with differing opinions, was higher than incivility in ideologically *homogeneous* political groups, Welch's t-test showed this difference is not significant (t = 1.545, df = 11, p = 0.15). A detailed monthly trend from 2015 to 2019, shown in **Figure 4B**, shows small yet growing fluctuations in the proportion of incivility in ideologically homogeneous and heterogeneous political groups. Two noticeable spikes in ideologically heterogeneous groups, which did not have corresponding spikes in the homogeneous groups, occurred in January 2017 and March 2018. The former spike overlapped with executive order 13,769 (also known as Muslim Ban) and Trump's inauguration. The spike in March 2018 could be attributable to the breaking news of Cambridge Analytica's involvement in Trump's presidential campaign. In turn, there were two spikes in incivility in ideologically homogeneous political groups that did not occur in ideological *heterogeneous* groups, in December 2018 and March 2019. The former can be linked to the longest US government shutdown in history and the latter was due to the release of the Mueller Report about Russian interference in the 2016 election.

Incivility in mixed groups that were ideologically heterogeneous spiked in 2010 and 2012 and reached its lowest levels in 2015. Sampling comments from mixed sub-reddits suggest that the reasons for these spikes were discussions about healthcare reform in 2010, whereas the peak in 2012 was due to the presidential election. In turn, incivility in mixed groups that were ideologically homogeneous fluctuated within a small range of 10–12%, significantly lower than that in mixed heterogeneous groups (t = 2.881, df = 11, p < 0.05). Furthermore, the monthly proportion of incivility in ideologically heterogeneous mixed groups from 2015 to 2019 was also significantly higher than that in homogeneous mixed groups (t = 9.439, df = 59, p < 0.001), indicating - again - greater usage of incivility in ostensibly non-political groups where both liberals and conservatives sometimes discuss politics.

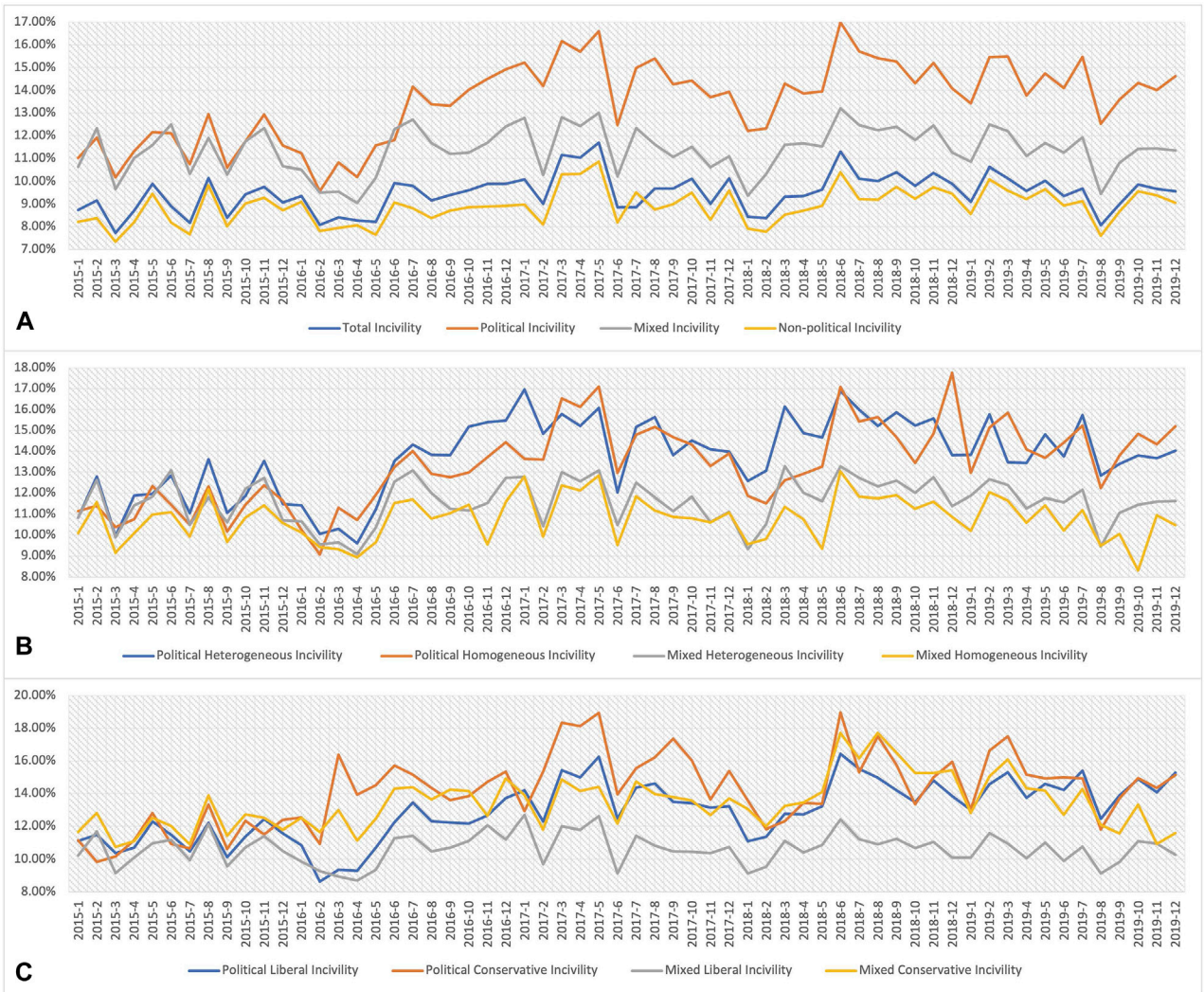

FIGURE 4 | Monthly Incivility Proportion from January 2015 to December 2019. Note. The percentage shown in the graph is the proportion of incivility content over relative total content. **(A)** is the proportion of incivility overall and in three main categories; **(B)** is the proportion of incivility in political heterogeneous/homogeneous and mixed heterogeneous/homogeneous groups; **(C)** is the proportion of incivility in political liberal/conservative and mixed liberal/conservative groups.

To answer RQ3, ideologically heterogeneous mixed groups entail more incivility than homogeneous groups, but it is not the case that discussions in ideologically heterogeneous political groups are necessarily more uncivil than discussions in ideologically homogeneous political groups. In addition, discussions in ideologically heterogeneous mixed groups were significantly less uncivil than those in ideologically heterogeneous political groups ($t = -12.987$, $df = 59$, $p < 0.001$), confirming recent findings (Rajadesingan et al., 2021).

Conservative vs Liberal Homogeneous Groups

Next, we examined the fluctuations in incivility in ideologically homogeneous, liberal or conservative political and mixed subreddits. The bars in **Figure 3** show that incivility was rather stable in liberal groups - especially mixed - as compared to conservative groups. Between 2008 to 2015, incivility in homogeneous liberal political groups gradually decreased from 14.98 to 11%, and then returned back to 14.97% in 2019. In turn, the proportion of incivility in homogeneous *liberal* groups that were mixed (discussing non-political issues but sometimes diverting to politics) decreased before 2015 and remained stable at around 11%, suggesting that the effects of the anti-harassment policy initiated by Reddit were especially effective in liberal mixed groups (perhaps because these groups were victims of disproportionate amount of trolling and harassment prior to the policy).

When it comes to incivility in ideologically homogeneous conservative groups, the oldest identified conservative subreddit was founded in 2009. Before 2015, the proportion of incivility in conservative political groups reached two peaks in 2011 and 2014. The peak in 2011 may be linked to the nomination

for the 2012 presidential election. In turn, uncivil comments from 2014 were mostly about gun control, minority groups, income equality, and climate issues, probably reacting to offline events such as legalization of same-sex marriage in several states, the raising of the minimum wage, and news about mass shootings and gun laws. A rapid increase in incivility in conservative political sub-reddits occurred after 2015 and gradually declined after 2017, consistent with the findings of political incivility during the 2016 election period (Nithyanand, et al., 2017a). In conservative mixed groups, incivility peaked in 2010 and declined to 9.35% in 2013 and then slowly climbed back to 15.58% in 2018. Conservative mixed groups were least affected by Reddit's anti-harassment policy, as incivility around 2015 was not at its lowest point.

To answer RQ 4, we see that before 2015, incivility in liberal political and mixed groups was higher than in conservative political and mixed groups. The trend reversed after 2015, with the highest incivility proportion in conservative political groups, followed by conservative mixed groups, liberal political groups, and liberal mixed groups. In fact, Welch's *t*-tests using monthly trends from 2015 to 2019, shown in **Figure 4C**, confirm that the proportion of incivility in conservative political groups was significantly higher than in liberal political groups (t = -6.304, df = 59, $p < 0.001$). Incivility in conservative mixed groups was also significantly higher than in liberal mixed groups (t = -16.049, df = 59, $p < 0.001$). Given that conservative news media are more likely to use outrage and divisive language (Sobieraj and Berry, 2011), this difference could be a reflection of the mainstream political discourse.[6]

Topics in Mixed and Non-political Groups

To answer RQ5, we turn to the non-political topic categories in mixed and non-political sub-reddits. Incivility in mixed discussions varied across topics and years (shown in **Figure 5A**).[7] Discussions about games, sports, and memes were most uncivil, perhaps because games and sports are ego-involving and, similarly to politics, generate an us-versus-them divide. In contrast, health and science/technology sub-reddits were the least uncivil, the former likely because many health sub-reddits discuss marijuana legitimization, which is supported by most participants, and the latter likely because most science and technology sub-reddits focus on problem-solving, which again, does not generate high incivility. When it comes to incivility in non-political sub-reddits, shown in **Figure 5B**, discussion about science and technology and pets and animals were most civil, whereas sports, memes, and entertainment generated more incivility on average.

Lastly, we calculated the average incivility proportion for all categories from 2008 to 2019, shown in **Figure 6** (See **Supplementary Appendix SF** for detailed statistics). Interestingly, the highest average proportion of incivility was found in the mixed gaming category, higher even than in political sub-reddits (Welch's *t*-test confirmed the significance of the difference, t = 4.758, df = 59, $p < 0.001$). Online gaming communities have their unique culture that often validates disparaging or disrespectful language, leading to this high aggregate proportion. Additionally reviewing comments from the mixed gaming sub-reddits suggests that those were mostly massive multiplayer online games with international servers that require a high level of communication among players and that are competitive by design, providing a hotbed for uncivil discourses.

External Events

The fluctuations of incivility in different categories were addressed throughout above, with some spikes in response to offline events and the changes in platform policies. Incivility in political and mixed groups tends to surge around highly contentious political events, such as election campaigns (e.g., 2016 presidential election), political scandals (e.g., Cambridge Analytica), and controversial orders (e.g., Musilm ban). Incivility in non-political groups also shows some spike during offline events such as sport events (e.g., 2019 super bowl), industrial scandals (e.g., 2016–2017 US gymnastic sexual abuse scandal), and industrial controversies (e.g., 2016 complaints about sexualized characters in games). Furthermore, there was a sharp decrease in incivility in all categories except mixed conservative groups in 2015, which corresponds to Reddit's anti-harassment policy. Consistent with the findings of previous research (Buntain et al., 2021), incivility in most categories after 2015 remained at a stable level, suggesting the intervention has both immediate and long-lasting effects.

DISCUSSION

Even though incivility is a growing concern for the public, politicians, and social media platforms, we know relatively little about its fluctuation's over-time and its prevalence across different types of online discussions. This study offers this key descriptive evidence, showing how incivility developed over-time on Reddit in political, mixed, and non-political groups, and also whether and how it differed in each group. We relied on a combination of machine learning methods and traditional statistical inference to examine the dynamics of online incivility on Reddit, an increasingly popular social media platform (Pew Research Center, 2021).

Our findings suggest that extant worries about the prevalence and rapid growth of online incivility may have been overstated. Incivility is not ubiquitous in Reddit discussions and has not dramatically grown in recent years. Its proportion is rather consistent, oscillating between 8 and 12%. The illusion of ever more incivility is due to the increasing volume of total online discussion in general, yet - again - the proportion of incivility to this overall volume of content is relatively stable. We also note that even though Reddit could invite greater incivility than

[6]When using the *yearly* data from 2008 to 2019, we find no significant differences in incivility between liberal and conservative groups, both political (t = 0.403, df = 10, p = 0.695) and mixed (t =-1.005, df = 10, p = 0.339).

[7]Although ten distinct categories were identified for *non-political* sub-reddits, only eight were found in *mixed* groups (and are used for the comparison, i.e., the category of Music and Pets/Animals was not present in mixed groups).

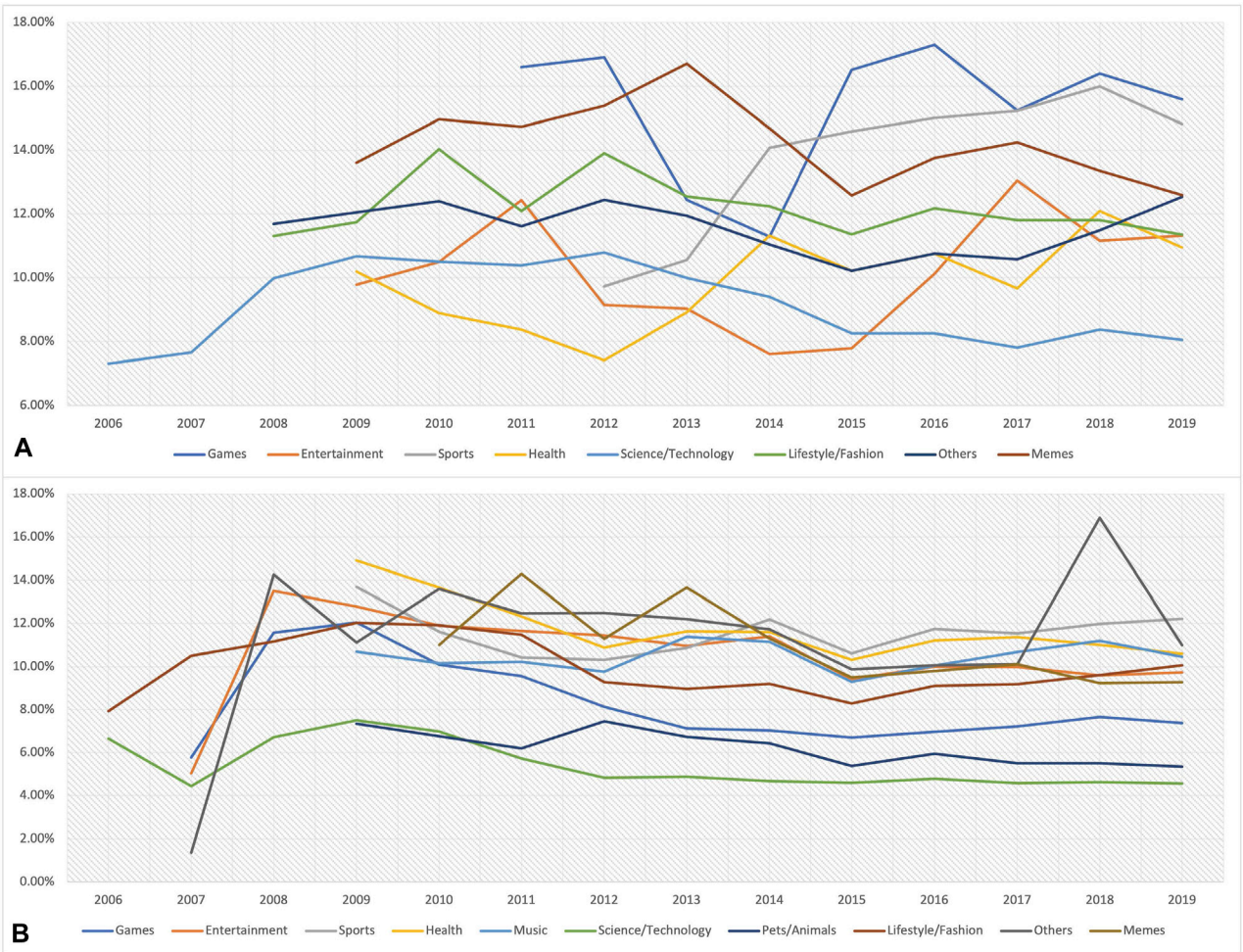

FIGURE 5 | Yearly Incivility Proportion in Mixed and Non-political Topics from 2006 to 2019. Note. The percentage shown in the graph is the proportion of incivility content over relative total content. **(A)** is incivility in mixed topics; and **(B)** is incivility in non-political topics.

Facebook or Twitter, due to its largely anonymous nature, the estimates in our data are largely similar to those from studies of other social media platforms (Siegel et al., 2018; Theocharis, et al., 2020).

Our other noteworthy findings relate to the differences in incivility across different categories of online discussion spaces. For one, consistent with anecdotal observations, users encounter more name-calling and disparaging or vulgar language in online discussions revolving around politics. That is, incivility is higher in political groups, followed by mixed groups whose focus is not politics, but which nevertheless entail socio-political discussions, and then non-political groups, where users discuss politics only rarely, if at all. A notable exception to this overall pattern is the mixed gaming category, where the aggregate proportion of incivility across all the years is higher than in political groups. Unlike other mixed groups, where incivility may be closely moderated and restricted by group members, gaming groups are known for endorsing incivility as a special social norm and encouraging uncivil behaviors such as flaming and trolling (Shen et al., 2020). Thus, incivility is likely to be promoted in such groups no matter whether discussions revolve around games or politics.

Second, even though ideologically diverse political discussions are seen as the breeding ground for uncivil discourse (Rossini et al., 2021), political sub-reddits involving participants expressing liberal and conservative perspectives are not necessarily more uncivil than ideologically homogeneous political groups. Furthermore, it is the ideologically heterogeneous mixed groups, where discussion about political issues may be unexpected and/or auxiliary and which involve diverse discussants, that entail *less* incivility than ideologically heterogeneous political groups, the sole purpose of which is to discuss politics. That is, heterogeneous political discourse is *less* uncivil in mixed sub-reddits than in political sub-reddits, consistent with the findings about uncivil cross-partisan discussions in non-political versus political online spaces (Rajadesingan, et al., 2021). It is possible that political

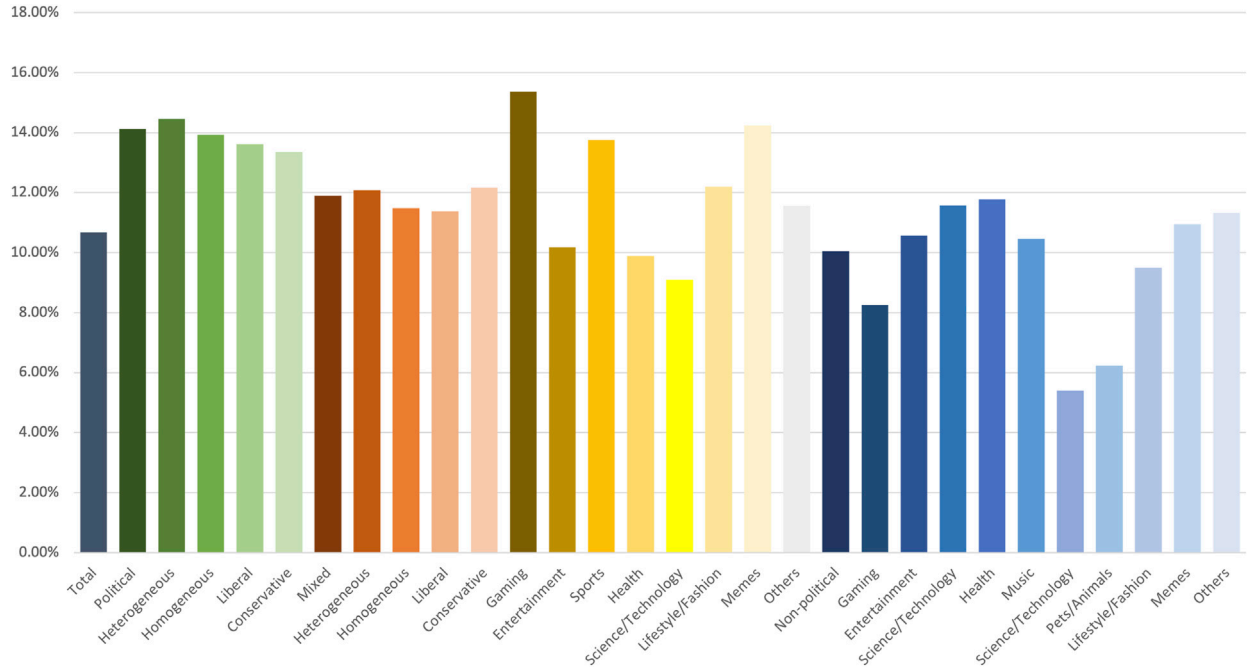

FIGURE 6 | Average yearly Incivility Proportion for All Categories from 2008 to 2019.

discourse is carefully moderated and restricted by moderators and members in sub-reddits that are designed for non-political topics, thereby preventing incivility. Alternatively, it may be the case that once users establish common ground on non-political topics (e.g., as chihuahua owners or Kardashians' fans), political disagreement with dissimilar discussants does not generate the same levels of emotional response, and thus incivility, as political disagreement in groups solely dedicated to current events and potentially divisive policies. Even though our large-scale project cannot speak to the underlying mechanisms, our findings clearly suggest that the dynamics of political discourse online are contingent on social context, such that differences in the types of conversation lead to different expressions of incivility.

Third, platform specific as well as exogenous factors may powerfully shape online discourse, trends in incivility included. With regard to the former, the presented patterns underscore the effectiveness of anti-harassment policies by social media platforms. In 2015, when Reddit allowed its users to report abuse and harassment and consequently banned sub-reddits promoting racism or anti-Semitism, overall incivility on the platform dramatically declined. In turn, underscoring the influence of the overall divisive political environment on online discussions in the subsequent years, we show that incivility clearly increased around the contentious 2016 elections and during Trump's presidency. Also, external socio-political events such as debates about welfare, gun control, or sexual minorities, also led to fluctuations in incivility, in line with previous research about political events impacting the temporal dynamics of incivility (Theocharis, et al., 2020). When these external events are divisive or controversial, peoples' expressions and exchanges on social media may get heated and uncivil.

In fact, we note that after 2015, incivility in political groups increased at higher rates than on the platform in the aggregate and that both political and mixed conservative groups generated significantly more incivility than liberal groups. This suggests differential effects of the political environment. It could be that conservatives were more susceptible to the polarized context, especially during the presidency of Donald Trump, known for his devise and often inappropriate rhetoric, which could have 'trickled down' to online communities on Reddit. In a related vein, this difference could be a result of conservatives consuming news from conservative sources known for their inflammatory expression (Sobieraj and Berry, 2011). Picking up the elite cues, either from politicians or news media, conservative Reddit users could be adopting certain expressions in their political discourse or using it as a basis for online discussion.

When interpreting these findings, a few limitations of our project should be kept in mind. First, future work should apply more sophisticated classification of political, non-political, and mixed groups, using machine learning applied at post level to automatically detect whether discourse is political only, non-political only, or both. Second, as we cannot access the deleted or removed comments, our results may be biased. Even though deleting and removing comments could be for various reasons, comments which are extremely uncivil were likely to be removed or deleted and so the observed cases of incivility may be milder, which may have led to underestimations of incivility on Reddit. Also, our findings may not generalize to other social media platforms, such as Facebook or YouTube. Reddit has a

unique culture and is known for its grassroots - as opposed to algorithmic - moderation system. As such, the uncivil discourse patterns observed on Reddit may not be found on other platforms. The fact that our estimates are largely similar to those detected on Twitter (Theocharis, et al., 2020), suggests certain robustness to our findings. Yet naturally, a systematic cross-platform work would be an important addition to the literature. Perhaps most importantly, our analysis only takes into account users' posts or comments, i.e., textual expressions. As such, we lose the information conveyed via memes, pictures, and videos.

Despite these limitations, our research is the first to offer systematic descriptive evidence of temporal dynamics of incivility on Reddit, over 11 years, across various categories of discussions, and focusing on thousands of sub-reddits that account for 95% of users and comments over this time period. Perhaps counter-intuitively, the rise in incivility has not been as steep as many observers fear and continues to constitute a similar fraction of the overall online discussion (so naturally increasing in total, but not proportionally), with some important variations across different contexts of the overall online public sphere. We hope that future work addresses these different dynamics and mechanisms, shedding more detailed light on the role of group culture, topical influence, offline socio-political events, platform level interventions, such as reporting or moderating systems, and the users themselves. All these macro, meso-, and micro-level factors influence incivility and need to be accounted for conceptually and analytically. Inasmuch as name-calling, disparaging and vulgar language, and other personal attacks have negative effects on public discourse, online discussions, social media users, and social media platforms themselves (Liang and Zhang, 2021), these investigations are important. As social media platforms have become a major source of news and information and an important channel for political discussion, understanding the complexity of online incivility is the necessary first step to promote a healthy dynamic of political deliberation in contemporary democracies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://pushshift.io/using-bigquery-with-reddit-data/.

AUTHOR CONTRIBUTIONS

QS: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing—Original Draft, Visualization MW: Conceptualization, Validation, Resources, Writing—Review and; Editing, Supervision SD: Methodology, Software, Data classification, Data Curation, Visualization, and Editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpos.2021.741605/full#supplementary-material

REFERENCES

Alexa (2019). Top Sites in United States. Available at: http://www.alexa.com/topsites/countries/US (Accessed December 14, 2020).

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., and Ladwig, P. (2014). The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies. *J. Comput-mediat Comm.* 19 (3), 373–387. doi:10.1111/jcc4.12009

Baer, D. (2021). The 31 Biggest Subreddits. Available at: https://blog.oneupapp.io/biggest-subreddits/ (Accessed June 25, 2021).

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., et al. (2019). Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social media Data. *Am. Polit. Sci. Rev.* 113 (4), 883–901. doi:10.1017/s0003055419000352

Bauman, S., Toomey, R. B., and Walker, J. L. (2013). Associations Among Bullying, Cyberbullying, and Suicide in High School Students. *J. adolescence* 36 (2), 341–350. doi:10.1016/j.adolescence.2012.12.001

Berry, J. M., and Sobieraj, S. (2013). *The Outrage Industry: Political Opinion media and the New Incivility*. Oxford University Press.

Blackwell, L., Dimond, J., Schoenebeck, S., and Lampe, C. (2017). Classification and its Consequences for Online Harassment. *Proc. ACM Hum.-Comput. Interact.* 1, 1–19. doi:10.1145/3134659

Brooks, D. J., and Geer, J. G. (2007). Beyond Negativity: The Effects of Incivility on the Electorate. *Am. J. Polit. Sci.* 51 (1), 1–16. doi:10.1111/j.1540-5907.2007.00233.x

Buntain, C., Bonneau, R., Nagler, J., and Tucker, J. A. (2021). YouTube Recommendations and Effects on Sharing across Online Social Platforms. *Proc. ACM Hum.-Comput. Interact.* 5 (CSCW1), 1–26. doi:10.1145/3449085

Coe, K., Kenski, K., and Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *J. Commun.* 64 (4), 658–679. doi:10.1111/jcom.12104

Curry, D. (2021). *Reddit Revenue and Usage Statistics*. (London) BusinessofApps. Available at: https://www.businessofapps.com/data/reddit-statistics/ (Accessed July 12, 2021).

Davidson, S., Sun, Q., and Wojcieszak, M. (2020). Developing a New Classifier for Automated Identification of Incivility in Social media," in *Proceedings of the Fourth Workshop on Online Abuse and Harms*, 95–101.

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). *Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Eady, G., Nagler, J., Guess, A., Zilinsky, J., and Tucker, J. A. (2019). How many People Live in Political Bubbles on Social media? Evidence from Linked Survey and Twitter Data. *Sage Open* 9 (1), 2158244019832705. doi:10.1177/2158244019832705

Feldman, L., Stroud, N. J., Bimber, B., and Wojcieszak, M. (2013). Assessing Selective Exposure in Experiments: The Implications of Different Methodological Choices. *Commun. Methods Measures* 7 (3-4), 172–194. doi:10.1080/19312458.2013.813923

Fletcher, R., and Nielsen, R. K. (2018). Are People Incidentally Exposed to News on Social media? A Comparative Analysis. *New Media Soc.* 20 (7), 2450–2468. doi:10.1177/1461444817724170

Gervais, B. T. (2017). More Than Mimicry? the Role of Anger in Uncivil Reactions to Elite Political Incivility. *Int. J. Public Opin. Res.* 29 (3), 384–405. doi:10.1093/ijpor/edw010

Gervais, B. T. (2015). Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-Based experiment. *J. Inf. Techn. Polit.* 12 (2), 167–185. doi:10.1080/19331681.2014.997416

Greenlee, J. (2014). *The Political Consequences of Motherhood*. Ann Arbor, MI: University of Michigan Press.

Groshek, J., and Cutino, C. (2016). Meaner on mobile: Incivility and Impoliteness in Communicating Contentious Politics on Sociotechnical Networks. *Soc. Media+ Soc.* 2 (4), 2056305116677137. doi:10.1177/2056305116677137

Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *Am. J. Polit. Sci.* doi:10.1111/ajps.12589

He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). June). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning." in IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE, 1322–1328.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annu. Rev. Polit. Sci.* 22, 129–146. doi:10.1146/annurev-polisci-051117-073034

Jacoby, W. G. (2018). Neither Liberal Nor Conservative: Ideological Innocence in the American Public by Donald R.Kinder and Nathan P.Kalmoe. Chicago, IL, University of Chicago Press, 2017. 224 Pp. Paper, $26.00. *Polit. Sci. Q.* 133 (4), 758–760. doi:10.1002/polq.12849

Kenski, K., Coe, K., and Rains, S. A. (2020). Perceptions of Uncivil Discourse Online: An Examination of Types and Predictors. *Commun. Res.* 47 (6), 795–814. doi:10.1177/0093650217699933

Kim, J. W., Guess, A., Nyhan, B., and Reifler, J. (2020). The Distorting Prism of Social Media: How Self-Selection and Exposure to Incivility Fuel Online Comment Toxicity. *J. Commun.*.

KRC Research (2018). Civility in America 2018: Civility at Work and in Our Public Squares. Available at: https://www.webershandwick.com/wp-content/uploads/2018/06/Civility-in-America-VII-FINAL.pdf (Accessed December 14, 2020).

Leurs, K., and Zimmer, M. (2017). *Platform Values: An Introduction to the# AoIR16 Special Issue*.

Liang, H., and Zhang, X. (2021). Partisan Bias of Perceived Incivility and its Political Consequences: Evidence from Survey Experiments in Hong Kong.*J. Commun* 71 (3), 357–379. doi:10.1093/joc/jqab008

Lin, Y. (2021). *10 Reddit Statistics Every Marketer Should Know in 2021*. Ottawa, ON: Oberlo. Available at: https://www.oberlo.com/blog/reddit-statistics (Accessed July 12, 2021).

Massanari, A. (2017). #Gamergate and the Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. *New Media Soc.* 19 (3), 329–346. doi:10.1177/1461444815608807

Moor, P. J., Heuvelman, A., and Verleur, R. (2010). Flaming on Youtube. *Comput. Hum. Behav.* 26 (6), 1536–1546. doi:10.1016/j.chb.2010.05.023

Muddiman, A. (2021). "Conservatives and Incivility," in *Conservative Political Communication* (Milton Park, United Kingdom: Routledge), 119–136. doi:10.4324/9781351187237-8

Murphy, S. C. (2004). 'Live in Your World, Play in Ours': The Spaces of Video Game Identity. *J. Vis. Cult.* 3 (2), 223–238. doi:10.1177/1470412904044801

Mutz, D. C., and Reeves, B. (2005). The New Videomalaise: Effects of Televised Incivility on Political Trust. *Am. Polit. Sci. Rev.* 99 (1), 1–15. doi:10.1017/s0003055405051452

Nithyanand, R., Schaffner, B., and Gill, P. (2017b). "Measuring Offensive Speech in Online Political Discourse," in *7th {USENIX} Workshop on Free and Open Communications on the Internet ({FOCI} 17)*.

Nithyanand, R., Schaffner, B., and Gill, P. (2017a). *Online Political Discourse in the Trump Era*. arXiv preprint. arXiv:1711.05303.

Papacharissi, Z. (2004). Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups. *New Media Soc.* 6, 259–283. doi:10.1177/1461444804041444

Pew Research Center (2021). Social Media Use in 2021. Available at: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ (Accessed June 25, 2021).

Prior, M. (2007). *Post-broadcast Democracy: How media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press.

Rains, S. A., Kenski, K., Coe, K., and Harwood, J. (2017). Incivility and Political Identity on the Internet: Intergroup Factors as Predictors of Incivility in Discussions of News Online. *J. Comput-mediat Comm.* 22 (4), 163–178. doi:10.1111/jcc4.12191

Rajadesingan, A., Budak, C., and Resnick, P. (2021). *Political Discussion Is Abundant in Non-political Subreddits (And Less Toxic)*. arXiv:2104.09560.

Ransbotham, S., Fichman, R. G., Gopal, R., and Gupta, A. (2016). Special Section Introduction-Ubiquitous IT and Digital Vulnerabilities. *Inf. Syst. Res.* 27 (4), 834–847. doi:10.1287/isre.2016.0683

Reader, B. (2012). Free Press vs. Free Speech? the Rhetoric of "Civility" in Regard to Anonymous Online Comments. *Journalism Mass Commun. Q.* 89 (3), 495–513. doi:10.1177/1077699012447923

Reddit (2021). Reddit Policy. Available at: https://www.reddithelp.com/hc/en-us/categories/360003246511-Privacy-Security September 2, 2021).

Rossini, P. (2020). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*. doi:10.1177/0093650220921314

Rossini, P., Maia, R., and Maia, R. C. (2021). Characterizing Disagreement in Online Political Talk: Examining Incivility and Opinion Expression on News Websites and Facebook in Brazil. *J. Deliberative Democracy* 17 (1). doi:10.16997/10.16997/jdd.967

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). *DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter*. arXiv preprint arXiv: 1910.01108.

Santana, A. D. (2014). Virtuous or Vitriolic. *Journalism Pract.* 8 (1), 18–33. doi:10.1080/17512786.2013.813194

Seregina, A., and Schouten, J. W. (2017). Resolving Identity Ambiguity through Transcending Fandom. *Consumption Markets Cult.* 20 (2), 107–130. doi:10.1080/10253866.2016.1189417

Shen, C., Sun, Q., Kim, T., Wolff, G., Ratan, R., and Williams, D. (2020). Viral Vitriol: Predictors and Contagion of Online Toxicity in World of Tanks. *Comput. Hum. Behav.* 108, 106343. doi:10.1016/j.chb.2020.106343

Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., et al. (2018). *Measuring the Prevalence of Online Hate Speech, with an Application to the 2016 US Election*.

Silver, A., and Andrey, J. (2019). Public Attention to Extreme Weather as Reflected by Social media Activity. *J. Contingencies Crisis Manag.* 27 (4), 346–358. doi:10.1111/1468-5973.12265

Sobieraj, S., and Berry, J. M. (2011). From Incivility to Outrage: Political Discourse in Blogs, Talk Radio, and cable News. *Polit. Commun.* 28, 19–41. doi:10.1080/10584609.2010.542360

Stevens, H. R., Acic, I., and Taylor, L. D. (2021). Uncivil Reactions to Sexual Assault Online: Linguistic Features of News Reports Predict Discourse Incivility. *Cyberpsychol. Behav. Soc. Netw.* doi:10.1089/cyber.2021.0075

Sun, C., Qiu, X., Xu, Y., and Huang, X. (2019). How to Fine-Tune BERT for Text Classification." in China National Conference on Chinese Computational Linguistics. Cham: Springer, 194–206. doi:10.1007/978-3-030-32381-3_16

Theocharis, Y., Barberá, P., Fazekas, Z., and Popa, S. A. (2020). The Dynamics of Political Incivility on Twitter. *Sage Open* 10 (2),2158244020919447. doi:10.1177/2158244020919447

Thorson, K., and Wells, C. (2016). Curated Flows: A Framework for Mapping media Exposure in the Digital Age. *Commun. Theor.* 26 (3), 309–328. doi:10.1111/comt.12087

Top, A. (2018). 500 Global Sites 2017. Available from: https://www.alexa.com/topsites.

Vale, L., and Fernandes, T. (2018). Social media and Sports: Driving Fan Engagement with Football Clubs on Facebook. *J. Strateg. Marketing* 26 (1), 37–55. doi:10.1080/0965254x.2017.1359655

Warner, B. R., and Neville-Shepard, R. (2014). Echoes of a Conspiracy: Birthers, Truthers, and the Cultivation of Extremism. *Commun. Q.* 62 (1), 1–17. doi:10.1080/01463373.2013.822407

Wikipedia contributors (2021c). "List of Most-Followed Facebook Pages," in *Wikipedia, the Free Encyclopedia*. Available at: https://en.wikipedia.org/w/index.php?title=List_of_most-followed_Facebook_pages&oldid=1041704712 September 13, 2021).

Wikipedia contributors (2021b). "List of Most-Followed Twitter Accounts," in *Wikipedia, the Free Encyclopedia*. Available at: https://en.wikipedia.org/w/index.php?title=List_of_most-followed_Twitter_accounts&oldid=1043863300 September 13, 2021).

Wikipedia contributors (2021a). "List of Most-Subscribed YouTube Channels," in *Wikipedia, the Free Encyclopedia*. Available at: https://en.wikipedia.org/w/

index.php?title=List_of_most-subscribed_YouTube_channels&oldid=1043463938 September 13, 2021).

Wojcieszak, M. E., and Mutz, D. C. (2009). Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement? *J. Commun.* 59 (1), 40–56. doi:10.1111/j.1460-2466.2008.01403.x

Wojcieszak, M., Menchen-Trevino, E., Lee, S., and Huang-Isherwood, W. *(forthcoming). No Polarization from Partisam News: Over-time Evidence from Trace Data*. The International Journal of Press/Politics.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.