Connective Effervescence and Streaming Chat During Political Debates

Abstract

Recent advancements in online streaming technologies have re-centered the audience as an important part of live broadcasts, including live political events. In fall 2020, each of the U.S. presidential and vice presidential debates were streamed on a number of online platforms that provided an integrated streaming chat where the public could comment in real-time alongside the live debate video. Viewers could simultaneously tune into what the candidates were saying and see what a sample of their peers thought about the candidates. This study examines large samples of comments made in social chat feeds during the livestreamed debates on the ABC News, NBC News, and Fox News Facebook pages to quantify key features associated with the quality of political discussion on these platforms. The results reveal that consistent with quasi-anonymous, constrained nature of the dynamic chat, the comments made are generally short, include a substantial degree of toxicity and insults, and differ significantly in their content across platforms. These findings underscore the importance of further study of online streaming chat as a new source of potential influence on political attitudes and behavior.

The media environment has undergone significant changes over the past several decades. One of the most important technological developments of the past 40 years remains the advent of cable news, which afforded the public a much wider choice of how much and what type of news to consume (Prior, 2007). Although there have continued to be major developments in the delivery of news across a number of digital desktop and mobile platforms, the bulk of news media consumption still takes the form of a live audio or video broadcast (Allen et al., 2020). Many people continue to prefer watching or listening to news in these longstanding ways (Mitchell, 2018).

Our study focuses on advancements that have changed the way people watch these broadcasts in ways that re-center the importance of the audience by integrating popular social media features into the traditional viewing experience. For example, during broadcast news or political events, people frequently take to social media as a "second screen" (Gil de Zúñiga et al., 2015) to see what others think as they watch live broadcasts. Even more recently, the advent of live video with *integrated* streaming chat is exploding in popularity on platforms dominated by younger generations. "Streaming chat" offers a viewing experience where the live video and real-time commentary are embedded on a screen together, encouraging viewers to immerse themselves in both sources at the same time. This technology continues to grow in popularity, particularly within the context of political events, as evidenced by its recent implementation by various platforms during the 2020 Presidential Election Debates. Each general election debate was simultaneously broadcast on network and cable television and livestreamed on online platforms, including Facebook, YouTube, and Twitch, where viewers could post real-time comments alongside the video feed.

Although many people, particularly younger people, spend considerable time streaming video content,[1] there has been little scholarly attention to the dynamics of political expression within livestream chats. **Crucially, we believe that the phenomenon-**

[1] A substantial majority of young adults prefer to use online streaming to watch TV (Pew, 2017).

logy of active participation in streaming chat is sufficiently different from that of "commenting" in the form of the temporally discrete "posts" that define the social web that existing theoretical approaches cannot be easily adapted. Instead, we propose that the concept of "connective effervesence" can explain the appeal of streaming chat: bringing the experience of belonging to and acting within a temporally contiguous *crowd* to the bandwidth-constrained realm of digital communication.

Empirically, we collect large samples of the comments in the streaming chat from three debates leading up to the 2020 U.S. Presidential Election, spanning September 29th to October 22, 2020, on livestreams from three major network Facebook pages: NBC News, ABC News, and Fox News. Our analysis quantifies the relative length of the nearly 90,000 comments in the sample, the frequency at which the debate participants were mentioned in the comments, and the toxicity of comments, to characterize key features of the chats and how they are similar or different across each of the three chat feeds and three debates.

Consistent with the "effervescent" nature of dynamic discussion during live events, we find that comments are generally extremely short, a feature associated with lower quality and non-deliberative discussion. In addition, there is substantial toxicity and frequent insults in the comments across platforms. The toxicity within the chat streams was especially pronounced for the first presidential debate between Donald Trump and Joe Biden on September 29, 2020. On the ABC News and NBC News platforms, more than 1 in 4 comments was detected as toxic in this debate, referring to rude and disrespectful remarks that could alienate people from participating in discussion. We also detected differences in the content of discussion between Fox News and the other two pages, ABC News and NBC News. While previous research has pointed to differences in debate commentary across channels as influencing post-debate attitudes

(Gross et al., 2019), these differences in the chat stream point to another mechanism– the audience– by which changing the (online) channel changes the viewing experience, potentially influencing what becomes salient in the minds of the audience as they form candidate evaluations or make decisions.

Recent research has shown how news media coverage can exert a causal influence on the way that audiences engage in public expression online (King et al., 2017). We argue that the public expression on social media is important in the construction of news media as it is experienced by other viewers. The integration of social chat into live broadcasts changes the potential scope of media effects on political behavior and political attitudes. Unlike many previous iterations and advancements in the study of media effects on political attitudes that focus on political and media elites, streaming chat centers the audience as a potential source of influence.

From Passive Consumer to Co-Creator

In the broadcast era, one major effect of media was priming (Iyengar and Kinder, 1987). Broadcasts that focused on certain subjects more heavily could have the effect of changing the relative weight that individuals place on that subject when evaluating political figures (Iyengar and Kinder, 1987). During this era, news anchors and the content of the news broadcast were the main influencers. Within the context of political debates, beyond the candidates, the main potential mechanism for influence was pre- or post-debate commentary from political analysts (Fridkin et al., 2008). The role of the viewer was that of passive consumption.

In contrast, in earlier media regimes, with theater, opera, and political speeches and debates taking place live, audiences were generally active and involved. The audience member was constantly bombarded with boos and cheers, shouted slogans or laughter. The mass media era of the 20th and early 21st century entailed the "pacification of the audience" (Napoli, 2011). People became accustomed to consuming broadcast media in their own homes, and their perception of the role of the audience changed. This spilled over to live audiences, who became much less active than previous live audiences (constrained, to some extent, by the technology of voice and image amplification that centered the action on stage).

Streaming chat represents a return to a more active audience, to a re-orientation of the experience of being an audience member.

While a lot of non-news consumption still takes place on social media (Allen et al., 2020), news and social media are increasingly entwined. In recent years, surveys have shown about two-thirds of Americans reported getting at least some of their news via social media (Shearer and Gottfried, 2017). In the era of the hybrid media system, the audience now has a role of co-creator and influencer (Chadwick, 2017). Comments made on social media posts by news consumers can influence what information from a news post or video is recalled and what information is weighed in our evaluations of political events and figures. In fact, at times, social media comments can stand out and become more salient than the content of news articles and videos themselves–creating an outsized influence on opinions. Anspach and Carlson (2018) show that the recall of information from these posts can be responsible for the propagation of misinformation when the content of the comments is inaccurate.

The nature of integrated streaming chat on social media only amplifies the importance of the commenters. Many online platforms are set up such that the visual field includes both the video and the chat box. To avoid seeing the comments as part of the broadcast, consumers have to opt out of this experience. In contrast, in many cases, to see static comments on a newspaper article, YouTube video, or Facebook post, individuals have to actively opt into clicking on this information below the post. As a less purposive activity, a wider range of consumers may be exposed to comments from streaming chats.

Just as scholars have dedicated substantial time in quantifying the agenda setting and priming roles of live broadcast news events, it is increasingly important to understand what is attracting attention in the comment sections of their livestreamed counterparts. Support for political figures may be subject to the whims of the sentiment of the commenters in the chat stream. Asbury et al. (2020) find that respondents encouraged to watch a 2019 Democratic primary debate with streaming chat on Facebook came away with more negative evaluations of specifically those candidates who received disproportionate criticism within the streaming chat.

Connective Effervescence

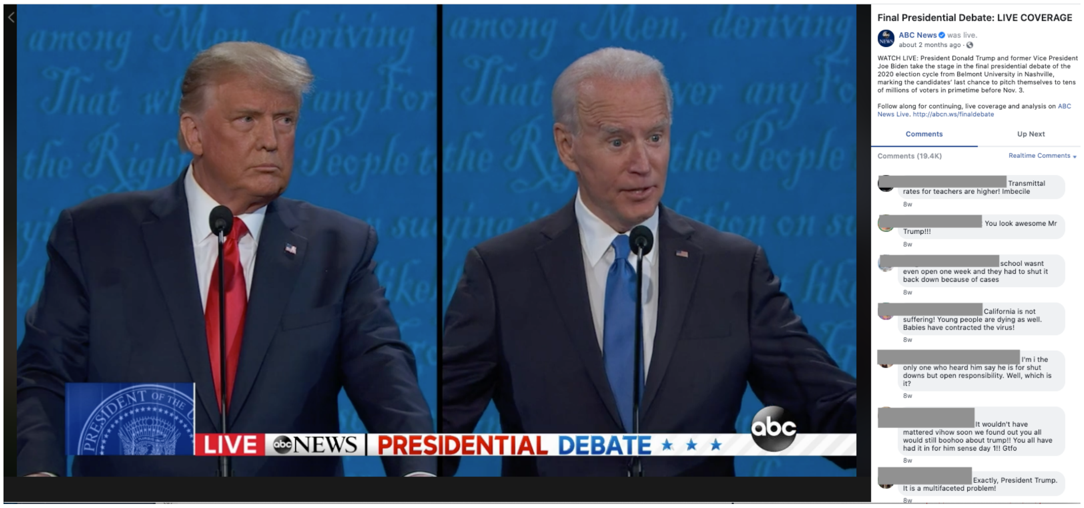
The format of the streaming chat also seriously constrains how people choose what to say. The primary differences between participating in a streaming chat and leaving comments in other online spaces is the *speed* of the chat. Protracted conversations are impossible; instead, commenters primarily react directly to what is happening in the online broadcast. The shared knowledge that the communal attention is focused on the video media makes it possible to speak without specifying a referent. Commenters can say "that was awesome" and rely on temporal proximity (or *contiguity*) for mutual context.[2]

This means that the already ferocious competition for attention in online platforms is only heightened. There are many stimuli happening simultaneously, and the main broadcast covers the majority of the viewers' screen. Figure 1 provides a static snapshot of this experience, but we recommend interested readers watch a few minutes of any popular recorded streaming chat for a better understanding of the dynamic experience.[3]

[2] This poses challenges for the analyst. We do not have sufficiently precise timestamp information on the comments to sync them up with the events of the broadcast.

[3] See A1 for links to streaming chats from this study.

Figure 1: Sample Screenshot of Presidential Debate Streamed on ABC News Facebook



The primary purpose for many commenters is thus better understood as *collective expression* than anything approaching persuasion, let alone deliberation. (Ford et al., 2017) **explains the phenomenology of streaming chat as "crowdspeak," and provides a helpful benchmark for the rate at which messages appear in a large-scale streaming chat: 1.75 per second, so that each is on the screen for only a few seconds.**

Another useful metaphor is that of a fan at a sporting event: there is an expectation of yelling or otherwise vocally responding to the focal action. The goal is partially to antagonize your opponents and hearten your allies, but more directly, to become part of living mass of people, to experience the digital analogue of **Durkheim's "collective effervescence" (Durkheim and Swain, 2008).**

Although previous scholars have applied the concept of collective effervescence to describe other forms of digital communication, we argue that the technological affordances of streaming chat make this a qualitatively different experience. Text-based digital media are constrained by the rhythms of writing; the latency is too high to approach the experience of being a live crowd of people. One-to-one (or small group) videos are low latency, but cannot numerically create the crowd experience; bottlenecks in information transmission like internet bandwidth, screen definition and audio quality mean that even our advanced communication technology cannot replicate live crowds.

Streaming chat around a focal broadcast accomplishes both low latency (frequent updates) and **the presence of many distinct** actors **in the commenting "crowd"** (hundreds of distinct text posts): short bursts of emotion that allow each commenter to perform their role as part of the crowd, thereby experiencing what we call "connective effervescence." Overall, the dynamic nature of participating in a live online conversation among the mass public has the tendency to make comments frequent, short, affective, and non-deliberative. **Through the crowd experience, streaming chats become social, even if the commenters are not directly interacting with each other (Haimson and Tang, 2017).**

A key element of Durkheim's theory of collective effervesence is the projection of the shared energy generated by close contact and unified action onto some symbol or ritual object. For most video live-streams with streaming chats, the focal object is the streamer herself: the community collectively experiences the streamer and each other as she navigates a video game, cementing the centrality of video gaming in the groups' self-image. In this context, there is no inherent conflict; the streamer and community are jointly reacting to the game events.

In contrast, the Presidential debates come with conflict built in, and with focal symbolic objects that can unite a community through affirmation of their preferred side, but also, particularly, negation of the opponent. "Affective polarization" between partisans has emerged as a defining concern over the past decade (Iyengar et al., 2012; Mason, 2018), but the vast majority of the change in partisan affect has been driven by decreased warmth towards

the outparty: "negative partisanship" is central to how partisan identities currently operate (Abramowitz and McCoy, 2019). In livestreams, commenters are often united through a shared emotional connection (Haimson and Tang, 2017) to the focal objects, leading naturally to expressions consistent with affective polarization, including potentially toxic and insulting comments negating the opponent, as discussed below.

Although that trend has been largely symmetric across parties, the context of the 2020 US Presidential Election was anything but symmetric. Both Republicans and Democrats were voting on Trump. In an August 2020 poll conducted by the Pew Research Center, 56% of Biden supporters said—as part of an open-response survey question—that the main reason they supported Biden was "He is not Trump" (Pew, 2020); in contrast, only 19% of Trump supporters selected anti-Bidenism as their main motivation. Trump is in fact the focal figure for both Republicans and Democrats, and that we would expect these nominally symmetric debates to provide asymmetric experiences for each type of partisan.

Although we argue that the experience of connective effervescence produced by participating in a livestream streaming chat can satisfy needs for sociality or community that other digital communication technologies cannot, their effects on political polarization are normatively troubling. Previous research has shown the constraints on discussion, influencing comment length, can have direct effects on the quality of political conversations. Notably, Jaidka et al. (2019) show that interactions with U.S. political figures on Twitter became more civil and polite and more likely to include constructive information-sharing after Twitter expanded its character limit for tweets. **The self-segregation inherent in the partisan streams we study[4] implies a context of**

[4] Haimson and Tang (2017) note that often people come across streams on Facebook due to already following the streamer's page. It is likely that many of the commenters in the streams we study are

increased partisan homogeneity relative to the free-for-all of Twitter, however. Insofar as polarization is a concern, in this context we would expect to find something like "confidence polarization" (or the "majority illusion") as partisans can experience the legions of people who agree with them overwhelm any outburst from the other side (Lerman et al., 2016; Ortoleva and Snowberg, 2015).

Partisan self-segregation is not the only form of audience heterogeneity. As important is the heterogeneity in the experience of the audience members as consumers and co-creators. The majority of viewers do not comment; there are between 100,000 and 200,000 comments on each of the debate streams, and many of these are from repeat commenters. Even if the experience of the commenters is purely performative or emotive, the non-commenting audience members may not fully understand the process that is creating the stream of text that appears in their field of vision.

Spirals of Toxicity

Online conversations on social media tend to include "toxic" or "uncivil" rhetoric under certain conditions. These are comments that are rude, disrespectful, or represent personal attacks, insults or threats.[5] Commenters are more likely to make uncivil or toxic comments if they see others doing the same (Cheng et al., 2017; Kim et al., 2020; Shmargad et al., 2020) or if their initial posts of this nature are rewarded through feedback of other commenters (Kim et al., 2020; Shmargad et al., 2020). Moreover, toxic comments may lead others–those who are disenchanted by toxicity–to withdraw from actively participating in the online conversations (Theocharis et al., 2016), reinforcing toxic descriptive norms of those engaged in discussion. Integrated *streaming chats*

people that already are familiar with these news outlets, creating an expectation that those following the Fox News stream might include more Republicans, for example.

[5]There is an extensive literature defining and exploring different types of non-deliberative online speech (Chen, 2017; Masullo Chen et al., 2019). These are important, but given our interest in the wave of text that comprises a streaming chat, we are less concerned about drawing these fine distinctions than in the aggregate experience.

on social media further intensify the conditions that lead to toxicity in political online discussions. This is because the fast-paced nature of dynamic chats make it even harder to identify any specific commenters– heightening the appearance of anonymity, which is known to only increase inflammatory rhetoric online (Mungeam, 2011).

The political comments to which viewers are exposed on social media are also often systematically different from the types of conversations they see about politics in daily life. Those who post comments on social media pages devoted to news and politics tend to be more partisan than the average American (Kim et al., 2020), creating conditions that lead to incivility and otherwise negative forms of political discussion.[6] Political debates inherently involve criticism of the opponent, making it particularly likely that the comment sections of social media will follow suit. **Furthermore, there is specific evidence of the possibility of "emotional contagion" on streaming chat, as the tone of chat messages can influence the tone of later chat messages (Guo and Fussell, 2020), potentially increasing the degree of partisan hostility expressed by the crowd throughout the stream.**

Expanding upon previous research, in our study, we quantify more specifically the degree to which political debate chat streams are toxic–and how toxicity differs by the actor mentioned in a comment (e.g., Trump vs. Biden) and platform in which the stream is occurring (e.g., Fox vs. ABC vs. NBC). While previous research has focused on the overall degree of toxicity on political discussion in online political discussions, we identify if particular political figures are more likely to receive toxic commentary in certain outlet livestreams than others.

Co-Creation and Moderation

While the commenting "crowd" has a large hand in shaping the nature and tone of discussion, many streaming chats also have features that allow

[6]These forms of discussion might be deemed "negative" normatively, due to their effects in reducing trust, for example (Mutz and Reeves, 2005).

for moderation. For example, in Facebook streaming chats, moderators can "turn on" features that filter out profanity, specify specific words and phrases to be banned, set a minimum character length for comments that are displayed in the discussion, and potentially limit which users can post (e.g., only followers) and how frequently they can post (e.g., only every 10 seconds).[7] In a study of similar moderation tools used on Twitch, Seering et al. (2017) find that pro-active moderation features can discourage spam attempts, as a form of anti-social behavior, but generally do not significantly encourage more pro-social behaviors in the comments.

With moderation, the crowd can be conceived as a co-creator along with the outlet that is controlling the moderation of the chat. Importantly, these moderation features are often not directly observed by onlookers of the streaming chat. As a result, at its most extreme, if a chat is heavily moderated with substantial constraints on length and use of specific language, the viewing public may come away with a false impression of the wisdom of a particular crowd, compounding the potential "majority illusion" concerns.

Moderation is certainly not absolute. Due to the fast-paced nature of massive streaming chats, the moderation tools are almost necessarily blunt. Commenters can quickly circumvent filters by using alternative words and phrases that communicate the same sentiment and message– however innocuous or toxic it may be. For example, in the streaming chats used in this sample, we find some evidence of filtering against profanity within certain platforms (Appendix Figure A15). However, these filters did not constrain commenters from the ability to communicate toxic comments in other ways.

The potential for moderation also poses additional challenges for analysis

[7]The current features for Facebook chat moderation are available here: https://www.facebook.com/business/help/631291454317308?id=648321075955172. Accessed March 2021.

given the difficulty in observing precisely what constraints have been applied to a streaming chat. In this study, we limit our inferences to describing what the crowd gets to see as opposed to everything that specific members in the crowd attempted to say. Gatekeeping is a central part of media communication, whether it is the broadcast itself, or the chat and comment sections (Scacco et al., 2015).

Streaming Chat Sample

We collected samples of comments made on Facebook integrated streaming chats during the live broadcasting of each of the 2020 U.S. Presidential general election debates taking place in fall 2020: the September 29, 2020 debate between Donald Trump and Joe Biden moderated by Chris Wallace and Fox News, the October 7, 2020 Vice Presidential debate between Michael Pence and Kamala Harris moderated by Susan Page of *USA Today*, and the October 22, 2020 debate between Donald Trump moderated by Kristen Welker of NBC. Figure 1 provides an example of what the integrated streaming chat looks like to a viewer on Facebook. Our goal is to analyze specific dimensions of these comments: who is most frequently mentioned in the comments (i.e., which political figures are attracting the most attention, which may relate to downstream priming effects), the length of the comments (i.e., understanding the extent to which comments reflect the connective effervescence, more so than lengthy deliberation), and the amount of toxicity in the environment.

Each of these debates was televised on major news networks and simultaneously streamed on several online platforms, including many Facebook pages of major news channels. We choose these events as the focus of our study given the massive audience and breadth of platforms simultaneously covering the event, allowing for a comparison across platforms.[8] Moreover, Facebook is among the most popular and widely used social media platforms. An early 2019 Pew Research Center survey found that 69% of U.S. adults use Facebook, with a majority of users visiting Facebook at least once a day (Perrin and Anderson, 2019).

We collect a sample of comments made in the chats of streams on the Fox News Facebook page, NBC News Facebook page, and ABC News Facebook page. **We chose these networks because these were the networks that were designated to host the three planned presidential debates (though the ABC presidential debate did not occur as planned), and include some variation in terms of the typical partisan and ideological leanings of the audience, with Fox News attracting a more Republican-leaning audience than the other two platforms (e.g., see Jurkowitz and Walker (2020)).** This allows for a comparison of the experience viewers receive when consuming the debate across channel platforms. While it is true that people can select into viewing particular streams and channels, research has shown that the channel can also exert its own independent effect on attitudes (e.g., Martin and Yurukoglu (2017)). For example, in a study of debate viewing in 2016, Gross et al. (2019) show that exposure to post-debate commentary on Fox News led viewers to become more positive in their evaluations of Donald Trump relative to exposure to post-debate commentary on MSNBC. Our analysis will help illuminate potential differences in the chat streams of these different channels, pointing to an additional potential mechanism by which "changing the channel" online can influence candidate evaluations. Table 1 displays the number of comments in the sample per event and

[8] Nielsen reports that more than 60 million watched the presidential debates https://www.nielsen.com/us/en/press-releases/2020/media-advisory-final-presidential-debate-of-2020-draws-63-million-viewers/. While Nielsen ratings do not include online social media viewing, it is likely to be substantial. In 2016, Facebook reported that the ABC News Facebook stream of the presidential debate between Trump and Clinton received 6.2 million live views https://www.facebook.com/formedia/blog/trends-facebook-live-and-news-publishers.

channel. **For additional details on data collection, sampling procedures, and a list of the URLs to these live streams, see the Appendix and Appendix Table A1.**

Table 1: Comment Sample Size by Debate and Facebook Page

Event	ABC News	NBC News	Fox News
Sep 29th, First Presidential Debate	9871	9908	9786
Oct 7th, Vice Presidential Debate	9851	9968	9825
Oct 22nd, Second Presidential Debate	9980	7971	12549

To gather the sample, we extracted public comments available through the manual search feature on the Facebook page's replay of the live event. While this does not allow us to collect all comments, we were able to collect a similar sample of comments across events and platforms.[9] Thus, while the comments described here may not be representative of all chat streams of political events on Facebook news pages, the comparisons made in the results are based on a consistent method of sampling, and thus, should provide reliable information about potential differences within-event between platforms, and within-platform between debates. To supplement the initial sample, we also gather a small set of the "most relevant" comments from each platform, as determined by Facebook's algorithm. Because sorting by "most relevant" is an easily accessible feature for watching videos on Facebook, we view this as another theoretically interesting sample to describe. These results are in the online appendix.

Results

We first summarize the nature of the comments by describing their length and the frequency at which each debate participant is mentioned in the comments. **Figure 2 shows the number of words per comment by platform and debate, and Appendix Figure A1 shows the number of characters per comment. There is**

[9]Our sample does not include comment replies, so we are not able to speak to the nature of specific information exchanges between commenters.

some evidence that moderation tools may have been activated to prevent extremely short comments. The distributions for the first debates across platforms tend to be right-shifted away from zero, as are those for the Fox second and third debates. Nonetheless, overall, consistent with the theoretical description of connective effervescence, comments on the chat streams are generally quite short. The average length of comments spanned from 8.20 words in length during the ABC Vice Presidential Debate stream to 18.61 words, on average, in the Fox first presidential debate stream. Table 2 summarizes average comment length by platform and event. Overall, comments on the Fox Facebook stream tended to be slightly longer than the other platforms across events. However, all events had short comment length on average, relative to other forms of political commentary during live events. For example, comments made to the Politico "expert" live commentary during the first presidential debate had an average length of 26.5 words, while comments made to the FiveThirtyEight live expert analysis had an average length of 49.44 words.

Not only are the comments short, but the speed of these comments are also quite high. During the first Presidential debate, there are around 0.427, 0.43, and 0.46 comments per second on ABC, Fox, and NBC respectively.[10]

[10]**Because we collect the main sample using the mobile version URLs, which does not provide the exact time of comments being made, we sample the videos from a fixed time period during the debate and count the number of comments by hand using the "realtime comments" tab under the desktop version URLs within each video.**

Figure 2: Distribution of Comment Length by Facebook Page (Number of Words)

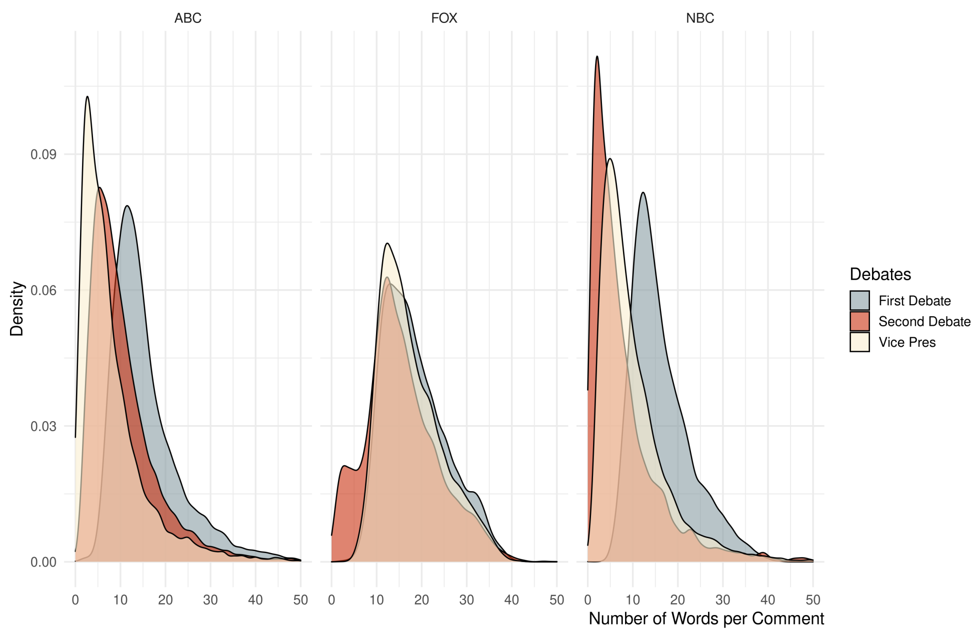


Table 2: Average Comment Length by Debate and Facebook Page

Event	ABC News	NBC News	Fox News
Sep 29th, First Presidential Debate	15.66	16.28	18.61
Oct 7th, Vice Presidential Debate	8.2	9.91	17.60
Oct 22nd, Second Presidential Debate	10.58	9.19	16.63

Note: Length refers to the number of words in a comment.

We next turn to identify which participants in the debates were most frequently mentioned in the comments. Here, we focus on direct mentions, as these would be the comments that the casual viewer of the online comments would be most likely to associate with the political actor. However, this may omit indirect comments made about a political figure in the moment.[11]

Figure 3 displays the proportion of comments that mentioned Trump, Biden, Pence, and Harris, respectively, in each stream. Three patterns emerge. First, a majority of

[11] There is no reason to believe that candidate names would be filtered out by content moderation. However, if moderation tools limit the visibility of extremely short comments, it is possible that streams where this is activated could prevent the visibility of short comments like "Trump 2020" or "Biden 2020." Our results describe the set of visible comments that the crowd can see.

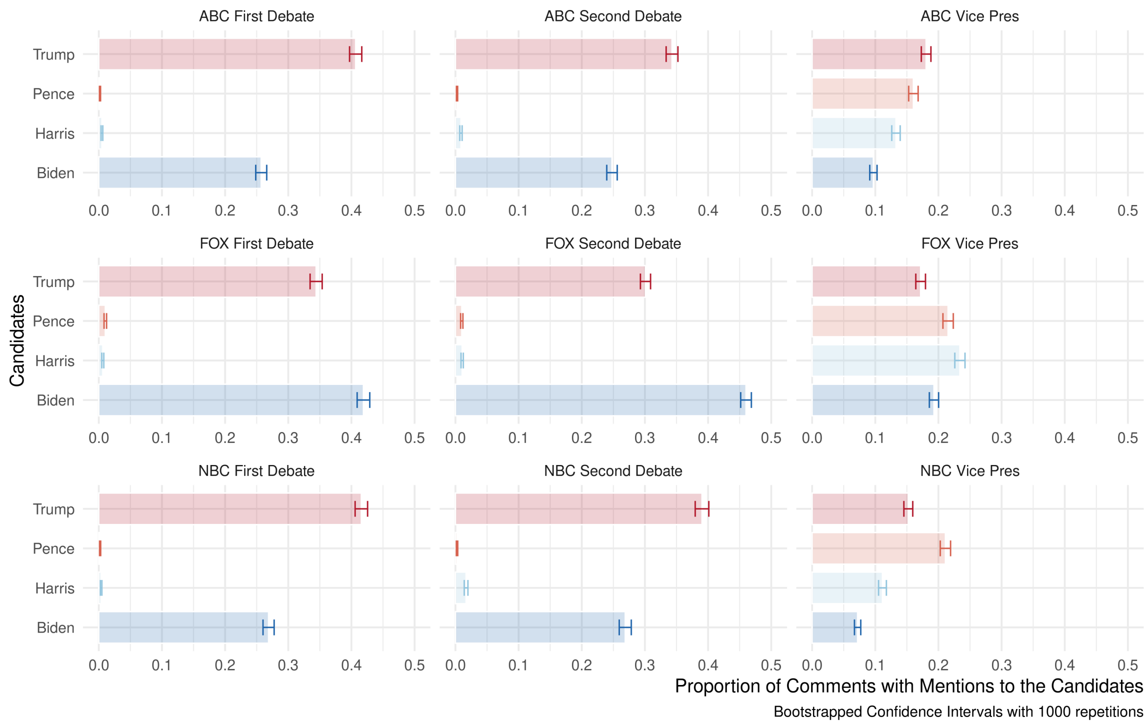
comments during the Presidential debates specifically reference at least one of the two participants.

Second, Pence and Harris are almost entirely omitted from conversations during the presidential debate, while in contrast, Trump and Biden are anything but absent in comments during the Vice Presidential debate. This concords with evidence that Vice Presidential candidates are rarely salient in the minds of voters. Here, it is only in the case of an event directly involving Harris and Pence, that their names become raised in online discussion about the election.

Third, Trump is consistently more likely to be mentioned than Biden in the ABC and NBC chat streams, across events. In contrast, Biden is slightly more likely to be mentioned than Trump in the Fox News streams of the presidential debates in the general sample of Facebook comments.[12] Given the partisan leanings of the audiences for the respective channels (pro-Biden for ABC and NBC, pro-Trump for FOX), **this may suggest** that commenters are more likely to mention the candidate they prefer less. Far from simply cheering for their side, commenters are unified in the experience of criticizing their opponents (Cheng et al., 2017; Kim et al., 2020; Shmargad et al., 2020). **That Trump is mentioned frequently across platforms may also reflect the centrality of Trump (as opposed to Biden) as a figure in the 2020 election.**

[12]These candidates are mentioned at similar rates in the sample of "most relevant" Facebook comments (Figure A8). It is unclear how Facebook's algorithm constructs this list, but it may well be explicitly trying to create partisan balance.

Figure 3: Proportion of Comments Mentioning a Candidate



Toxicity

To measure toxicity, we employ the Google Perspective API, which uses a machine learning model to score comments based on the likelihood someone would see it as "toxic": "a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion"[13] The API has previously been used in research to classify the comments made on YouTube videos (Obadimu et al., 2019), assisting in the classification of incivility in political discussions on Twitter (Theocharis et al., 2020), and detecting toxicity in comments on public Facebook pages of news outlets (Kim et al., 2020) and in political communities on Reddit (Rajadesingan et al., 2020).[14] In addition to toxicity,

[13]This definition is taken from the API documentation `https://support.perspectiveapi.com/s/about-the-api-key-concepts`. The API draws on a model based on millions of comments classified by human annotators.

[14]While any automated classifier is imperfect, Rajadesingan et al. (2020) finds that the Perspective API generally outperforms a single human labeller, providing a combination of accuracy and efficiency.

the model also detects a number of other attributes of comments, such as subdimensions of toxicity, including severe toxicity, threat, and insults. **Distributions of each measure are in Appendix Figure A2.**[15]

The machine learning models provide each comment with a score from 0 to 1 corresponding to the probability the comment has the given attribute. We dichotomize this score, and consider all comments with a score above .5 as having the attribute and below .5 as not having the attribute [16] Thus, when the results report the proportion of toxic comments, this reflects the proportion of comments above .5 on the toxicity score. Figure 4 displays the proportion of comments in the debate chat streams classified with a given attribute. **Appendix Table A2 and Table A3 provide examples of comments by their levels of each attribute**.

It is possible that moderation tools could affect the levels of toxicity observed in the comments, such as by banning certain words and phrases or activating the platform's profanity filter. In the appendix, we describe evidence on the possibility of this type of comment censoring in Appendix Figure A12 and Figure A15. We find some evidence that ABC and Fox had stronger filters against profanity than NBC, but there is little evidence that the moderation tools guarded against toxicity, overall.

Turning to the results, for the first presidential debate, more than 1 in 4 comments on the ABC News and NBC News chat streams were toxic, and nearly 30% were detected as

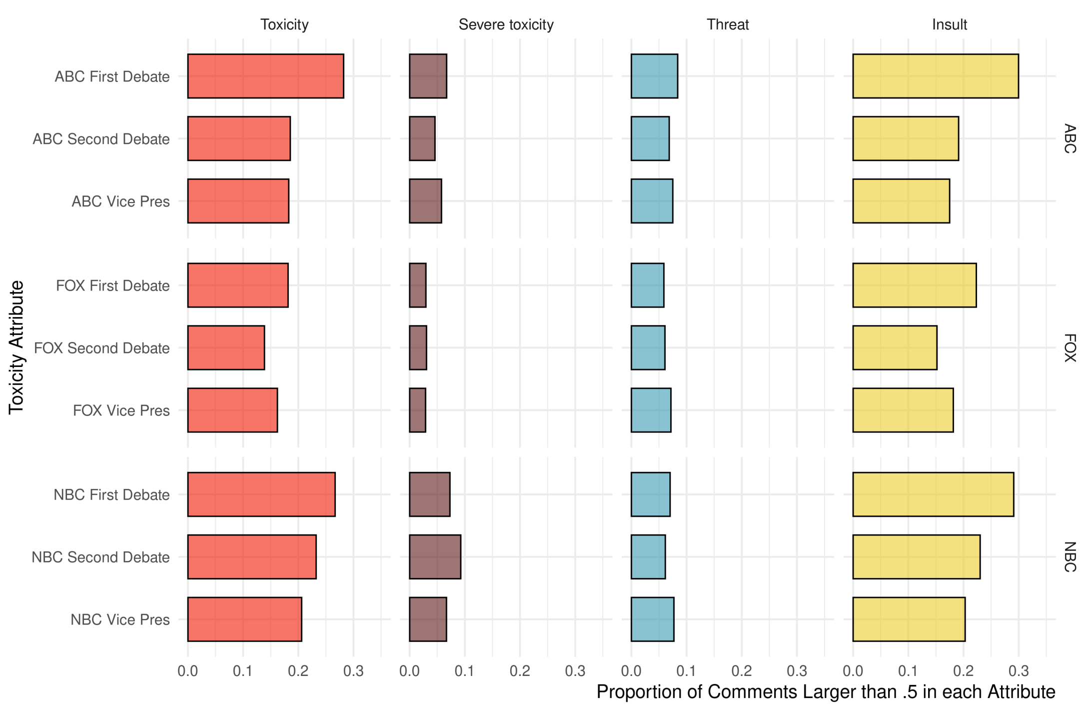
[15]According to the API documentation, severe toxicity is "A very hateful, aggressive, disrespectful comment or otherwise very likely to make a user leave a discussion or give up on sharing their perspective." Insults are, "Insulting, inflammatory, or negative comment towards a person or a group of people." Threat "Describes an intention to inflict pain, injury, or violence against an individual or group." Definitions are available here https://support.perspectiveapi.com/s/about-the-api-attributes-and-languages.

[16]The decision to dichotomize the toxicity scores comes intuitively from the nature of the measure. The API Perspective is built as a classification algorithm, where the score is the likelihood of a comment being classified as "toxic", and not a continuous measure. Therefore, our decision to work with outcome variable that replicates the intuition of the classification algorithm, whereby scores above .5 reflect an estimated majority of people perceiving the comment as toxic.

being insults. This degree of toxicity declined slightly for subsequent debates. Toxicity was overall slightly lower for the presidential debates within the Fox News chat stream; still, more than 15% and more than 20% of comments were detected as toxic or insults, respectively, in the first debate.

This pattern makes sense, **given that the candidates provide the focal objects to which commenters are reacting**: the first Presidential debate was widely considered the worst Presidential debate ever broadcast. Trump was unprecedentedly aggressive, frequently interrupting Biden and blithely refusing to answer questions posed directly to him by the moderator. Former debate moderators referred to it as "a hot mess inside a dumpster fire inside a train wreck" and "a shitshow;" Republican strategist Karl Rove summarized that "that was not very edifying or enlightening for the viewer" (Emmrich, 2020). The most memorable moment was when a frustrated Biden said to Trump, "Will you shut up, man?" after being continuously interrupted.

Figure 4: Proportion of Comments that are Toxic or Insulting by Debate and Platform



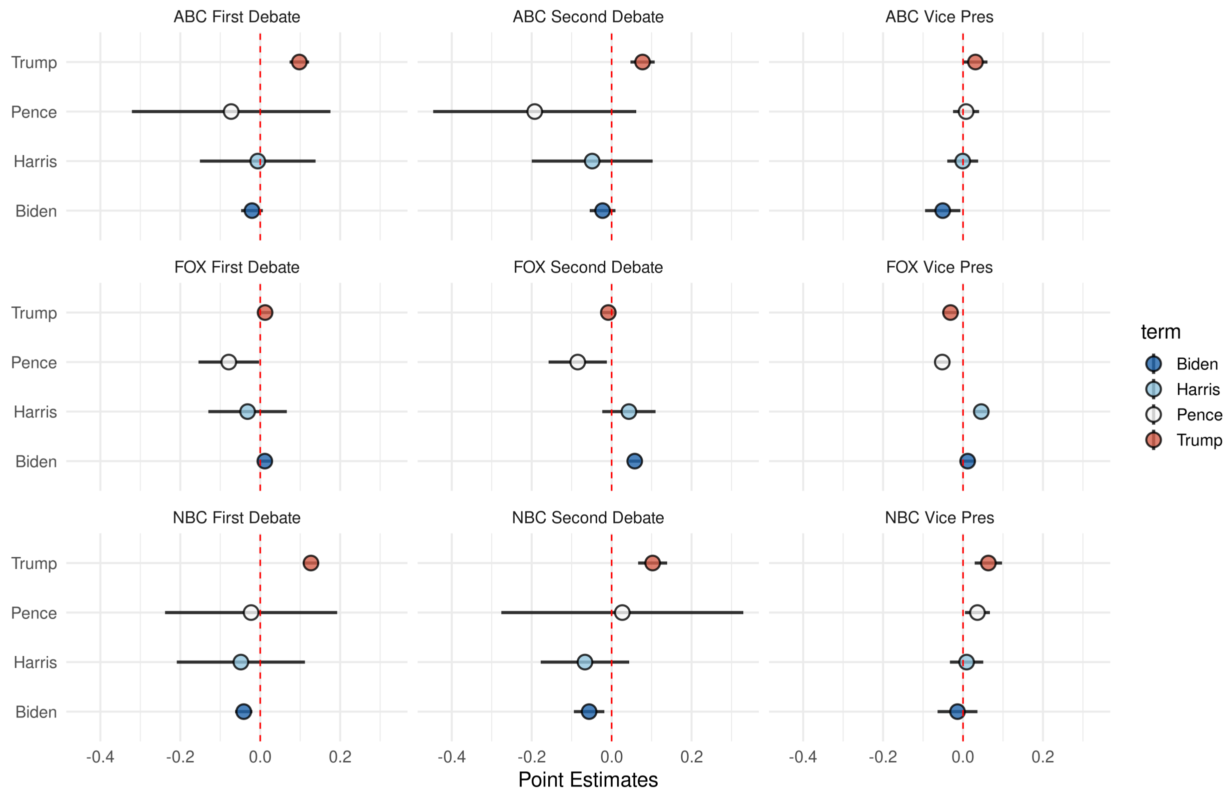
Though toxicity and insults are prevalent in the comments, remaining above 10% (toxicity) and 15% (insults) across platforms and debates, detected instances of severe toxicity or explicit threats of physical harm or violence were less common, overall.

The analysis next turns to quantifying the degree of toxicity in comments that mention a particular political actor: Trump, Biden, Harris, or Pence. Figure 5 displays the results of a linear probability model predicting the likelihood that a comment is toxic, if it mentions a particular actor, relative to comments that do not mention the actor. Regression results that include covariates for the number of likes on the comment and comment length are available in the appendix (Table A2). Figure A5 in the online appendix displays the average toxicity of comments that mention particular actors, by debate event and platform.

The results show that in the chat streams of ABC News and NBC News, comments that mention Trump are significantly more likely to be toxic, and in particular, relative to comments that mention Biden. In the presidential debates on Fox News, the results are less strikingly different between candidates, with Trump and Biden comments having similar levels of toxicity in the first debate and Trump receiving slightly less toxicity in the other debates. **Appendix Figure A4 displays the point estimates for the differences (and 95% confidence intervals) in toxicity between Trump vs. Biden and Harris vs. Pence for each debate and platform.**

For Harris and Pence, we focus primarily on the Vice Presidential debate, where these individuals were mentioned in a greater degree of comments. Here, the results differ somewhat substantially by platform. Comments that mention Republican Pence are just slightly more toxic, on average, than comments that mention Democrat Harris on ABC and NBC pages (Figure 5). On Fox News, in contrast, the pattern is reversed. In particular, comments that mention Harris are significantly more likely to be toxic than comments that mention Pence or Trump.

Figure 5: Predicting Toxic Comments, by Debate and Platform



In addition to specific candidate mentions, we find evidence of other features that are associated with an increase in the likelihood that a comment is toxic (Online Appendix Table A3). In particular, longer comments are significantly associated with toxicity in 8 of 9 of the channel-debate regression models. Again, this may point to the self-selective nature of comment generation in online political discussion. Even though there is evidence that creating the conditions for lengthier exchanges reduces negative or non-constructive discussion behaviors (Jaidka et al., 2019), within a given discussion context, longer comments may not equate to more constructive comments. In addition, toxic comments seem to be somewhat attention grabbing. The number of reactions to a post on Facebook is significantly and positively associated with toxicity in 6 of 9 regression models. This finding underscores the potential troubling dynamics of toxicity on online chat streams, whereby toxic behaviors may be rewarded, inducing even more toxicity (Cheng et al., 2017; Kim et al., 2020; Shmargad et al., 2020).

The association between toxic language and comments targeting particular candidates **suggests** that streaming chats are dominated by expressions of affective polarization, resembling closely the political environment in the recent years in American politics (Mason, 2018). Users watching the debates in their more ideologically-aligned news channel **may**use the anarchic and open streaming chats to vocalize their anger and disgust against their political opponent, exposing millions of other users to normatively concerning levels of toxicity.

In the online appendix, we provide several robustness checks for the main results discussed here. We report results using: bivariate analysis between toxicity and candidates mentions (Figure A3 and Table A5); statistical models using other content classifications retrieved from the Perspective API (Severe Toxicity, Insult and Threats, Figure A7); and using alternative cutoffs for the toxicity scores (Figure A6). Our robustness checks overwhelmingly go in the same direction of the trends highlighted in the paper.

Conclusion

Our results show that the comments in the political debate chat streams on Facebook tend to be short and contain a considerable degree of toxicity or insults– two features that have been systematically linked to normatively lower quality, less deliberative, and more polarizing discussion in existing research. These results reveal a tension between the potential for connective effervescence to increase a sense of sociality or engagement by providing new ways for the audience to become actively involved in politics and express themselves during elections, and the relative low quality and negative forms of discussion that are often generated when the conditions for discussion are quasi-anonymous and incentivize fast-paced, attention-grabbing remarks.

We also detect significant differences in the content of streams between networks. In particular, comments on ABC News and NBC News are more likely to mention Trump directly than to mention Biden, a pattern that is not observed for Fox News, where comments are more balanced between candidates in the presidential debates. The content differences also extend to measures of toxicity, whereby ABC News and NBC News tend to have somewhat higher proportions of toxic comments in the chat stream, and comments mentioning Trump specifically are more likely to be toxic on these platforms than other comments or comments that reference Biden. In contrast, Fox News shows fewer differences in the relative rates of toxicity between Trump and Biden in the presidential debates, and within the Vice Presidential debate, comments on the Fox News chat stream that mentioned Harris were significantly more likely to be toxic than comments that mentioned Pence, patterns that were not observed for ABC News or NBC News.

These content and differences in toxicity between the network news pages point to streaming chat as potentially another important mechanism by which media may influence the viewing experience, and subsequently affect several attitudinal and behavioral outcomes, such as candidate evaluation or willingness to participate in future political activities. While we cannot observe these behaviors in this study, they provide natural questions for future research on this new and understudied form of political engagement online.

More generally, this research design and analysis describes a prominent example of a broader phenomenon whereby the advent of livestreaming chat is fundamentally altering political news consumption. The general election debates are a widely watched political spectacle, drawing a large audience

to the online livestreaming platforms. In future research on streaming chat, the type of event, platform, and associated audience may shape the specific findings observed. For example, in a widely watched political event, the streaming comments are incredibly fast-paced. If a platform or a particular livestream on a platform has only a handful of viewers, the nature of comments may fall short of "connective effervescence." Instead, commenters may be in a space where they can have more deliberatve conversations with other viewers.

We, therefore, cannot definitively say that the differences observed in the Fox and ABC NBC News samples for the general election debates would generalize to all political events held on the platforms. We also believe that we might find even larger differences between networks if we had included Facebook pages from platforms to the more extreme left or right. Lastly, the findings also may differ depending on the moderation policies of a given platform. If moderators of a specific platform or of a specific livestreamed event on a given platform create policies on the type of and frequency of content a user can post, then this may also fundamentally alter the nature of the experience of streaming chat and what the crowd sees.

One limitation of this analysis is that we focus strictly on comments made on Facebook chat streams. Comments made on Facebook may be qualitatively different from the messages communicated elsewhere on social media, such as on Twitch, for example, another platform that relies heavily on integrated streaming chat, or on YouTube, which may have a different user base.

In addition, we might also see a difference between "livestreams" of broadcast media with streaming chat (the focus of this study) and livestreams by self-identified "streamers." The former is more similar to the experience of "dual-screening," where the audience of a traditional media broadcast are simultaneously discussing the proceedings on an alternative device and platform (canonically, looking at Twitter on their smartphone while watching a broadcast on their television). "Streamers," in contrast, are producing video explicitly for their streaming audience, who are chatting with each other *and the streamer* in an integrated online platform like Twitch. The two styles are often integrated, where a streamer and her audience watch, say, a live sports broadcast and interact with each other, with the streamer serving as the focal figure and interacting with the chat.

These streams have also started to intersect with mainstream politics. For example, Congresswoman Alexandria Ocasio-Cortez, a frequent first Congressional adopter of new media technology, provided a degree of mainstream legitimacy to already popular video game livestreamers and introduced streaming video technology to traditional media consumers when she participated in a livestreamed game of "Among US" in late 2020. Joined by Congresswoman Ilhan Omar and streaming luminaries like Pokimane, HasanAbi, Disguised Toast and DrLupo, the livestream reached over 430,000 concurrent viewers, making it one of the top ten streams to date (Kastrenakes, 2020). One of the co-streamers, HasanAbi, regularly streams political content on his channel.

All told, this technology is far from settled; we expect that there will continue to be significant innovation in both the platform affordances of streaming chat and the ways in which streamers and their audiences choose to use them. The possibility of "connective effervescence" that we theorize is an important step forward for digital communication technology, pushing the boundaries of social experiences. Our study quantifies key features of these chat streams and suggests that providing more pathways for political expression is not always better and may foster the types of toxicity that reflects and reinforce the polarized environment in American politics.

References

Abramowitz, A. and McCoy, J. (2019). United states: Racial resentment, negative partisanship, and polarization in trumps america. *The ANNALS of the American Academy of Political and Social Science*, 681(1):137–156.

Allen, J., Howland, B., Mobius, M., Rothschild, D., and Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14):eaay3539.

Anspach, N. M. and Carlson, T. N. (2018). What to believe? social media commentary and belief in misinformation. *Political Behavior*, pages 1–22.

Asbury, V., Chang, K.-C., McCabe, K., Munger, K., and Ventura, T. (2020). The effect of streaming chat on perceptions of debates. https://osf.io/5ux4s.

Chadwick, A. (2017). *The hybrid media system: Politics and power*. Oxford University Press.

Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. Springer.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., and Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, page 12171230.

Durkheim, E. and Swain, J. W. (2008). *The elementary forms of the religious life*. Courier Corporation.

Emmrich, S. (2020). Last nights presidential debate was a shitshow and a train wreck. can we survive two more like this? *Vogue*, September 30, 2020.

Ford, C., Gardner, D., Horgan, L. E., Liu, C., Tsaasan, A., Nardi, B., and Rickman, J. (2017). Chat speed op pogchamp: Practices of coherence in massive twitch chat. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*, pages 858–871.

Fridkin, K. L., Kenney, P. J., Gershon, S. A., and Serignese Woodall, G. (2008). Spinning debates: The impact of the news media's coverage of the final 2004 presidential debate. *The International Journal of Press/Politics*, 13(1):29–51.

Gil de Zúñiga, H., Garcia-Perdomo, V., and McGregor, S. C. (2015). What is second screening? exploring motivations of second screen use and its effect on online political participation. *Journal of Communication*, 65(5):793–815.

Gross, K., Porter, E., and Wood, T. J. (2019). Identifying media effects through low-cost, multiwave field experiments. *Political Communication*, 36(2):272–287.

Guo, J. and Fussell, S. R. (2020). A preliminary study of emotional contagion in live streaming. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, pages 263–268.

Haimson, O. L. and Tang, J. C. (2017). What makes live events engaging on facebook live, periscope, and snapchat. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 48–60.

Iyengar, S. and Kinder, D. R. (1987). News that matters: Agenda-setting and priming in a television age. *News that Matters: Agenda-Setting and Priming in a Television Age*.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideology a social identity perspective on polarization. *Public opinion quarterly*, 76(3):405–431.

Jaidka, K., Zhou, A., and Lelkes, Y. (2019). Brevity is the soul of twitter: The constraint affordance and political discussion. *Journal of Communication*, 69(4):345–372.

Jurkowitz, Mark, A. M. E. S. and Walker, M. (2020). U.s. media polarization and the 2020 election: A nation divided. *Pew Research*.

Kastrenakes, J. (2020). Aoc's debut twitch stream is one of the biggest ever. *The Verge*. https://www.theverge.com/2020/10/20/21526164/aoc-among-us-twitch-debut-top-concurrent-viewers.

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. https://www.dartmouth.edu/~nyhan/comment-toxicity.pdf.

King, G., Schneer, B., and White, A. (2017). How the news media activate public expression and influence national agendas. *Science*, 358(6364):776–780.

Lerman, K., Yan, X., and Wu, X.-Z. (2016). The" majority illusion" in social networks. *PloS one*, 11(2):e0147617.

Martin, G. J. and Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99.

Mason, L. (2018). *Uncivil agreement: How politics became our identity.* University of Chicago Press.

Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., and Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media+ Society*, 5(3):2056305119862641.

Mitchell, A. (2018). Americans still prefer watching to reading the news and mostly still through television. *Pew Re-*

search Center. https://www.journalism.org/2018/12/03/americans-still-prefer-watching-to-reading-the-news-and-mostly-still-through-televi

Mungeam, F. (2011). Commenting on the news: How the degree of anonymity affects flaming online. Thesis, MA Program in Communication and Leadership Studies, Gonzaga University.

Mutz, D. C. and Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, pages 1–15.

Napoli, P. M. (2011). *Audience evolution: New technologies and the transformation of media audiences*. Columbia University Press.

Obadimu, A., Mead, E., Hussain, M. N., and Agarwal, N. (2019). Identifying toxicity within youtube video comment. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 214–223. Springer.

Ortoleva, P. and Snowberg, E. (2015). Overconfidence in political behavior. *American Economic Review*, 105(2):504–35.

Perrin, A. and Anderson, M. (2019). Share of u.s. adults using social media, including facebook, is mostly unchanged since 2018. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since

Pew (2017). About 6 in 10 young adults in u.s. primarily use online streaming to watch tv. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2017/09/13/about-6-in-10-young-adults-in-u-s-primarily-use-online-streaming-to-watch-tv/.

Pew (2020). Election 2020: Voters are highly engaged, but nearly half expect to have difficulties voting. *Pew Research Center*. https://www.pewresearch.org/politics/2020/08/13/election-2020-voters-are-highly-engaged-but-nearly-half-expect-to-have-difficulties

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Rajadesingan, A., Resnick, P., and Budak, C. (2020). Quick, community-specific learning: How distinctive toxicity norms are maintained in political subreddits. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 557–568.

Rinker, T. W. (2019). *sentimentr: Calculate Text Polarity Sentiment*. Buffalo, New York. version 2.7.1.

Scacco, J., Curry, A. L., and Stroud, N. J. (2015). Digital divisions: Organizational gatekeeping practices in the context of online news. *Editors Note*, page 106.

Seering, J., Kraut, R., and Dabbish, L. (2017). Shaping pro and anti-social behavior on twitch through moderation and example-setting. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 111–125.

Shearer, E. and Gottfried, J. (2017). News use across social media platforms 2017. *Pew Research Center*. https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/.

Shmargad, Y., Coe, K., Kenski, K., and Rains, S. (2020). Social norms and the dynamics of online incivility. http://www.yotamshmargad.com/wp-content/uploads/2020/11/Incivility-Norms.pdf.

Theocharis, Y., Barberá, P., Fazekas, Z., and Popa, S. A. (2020). The dynamics of political incivility on twitter. *Sage Open*, 10(2):2158244020919447.

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., and Parnet, O. (2016). A bad workman blames his tweets: the consequences of citizens' uncivil twitter use when interacting with party candidates. *Journal of communication*, 66(6):1007–1031.

Titiunik, R., Calonico, S., and Cattaneo, M. D. (2015). rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. *R Journal*.

Connective Effervescence and Streaming Chat During Political Debates

Online Appendix

Data Collection and Sampling Strategy

Our empirical analysis in based on a novel sample of Facebook comments made by users on streaming chats during the live broadcasting of three 2020 U.S presidential and Vice Presidential debates. In the main paper, we report results from the collection of unfiltered comments, and in the appendix, we report results from the comments filtered by Facebook as "most relevant." Both samples were collected using web-scrapping techniques. In this appendix, we describe in detail our methodological strategy to assemble our dataset of Facebook comments.

Table A1: URLs of Debate Live Streams

Livestream	Desktop URL	Mobile URL
ABC First Debate	https://www.facebook.com/86680728811/videos/335718177646214	https://m.facebook.com/story.php?story_fbid=335718177646214&id=86680728811
ABC Second Debate	https://www.facebook.com/86680728811/videos/407053660475705	https://m.facebook.com/story.php?story_fbid=407053660475705&id=86680728811
ABC Vice Debate	https://www.facebook.com/86680728811/videos/831144484089448	https://m.facebook.com/story.php?story_fbid=831144484089448&id=86680728811
Fox First Debate	https://www.facebook.com/15704546335/videos/365638327905155	https://m.facebook.com/story.php?story_fbid=365638327905155&id=15704546335
Fox Second Debate	https://www.facebook.com/15704546335/videos/805515450261042	https://m.facebook.com/story.php?story_fbid=805515450261042&id=15704546335
Fox Vice Debate	https://www.facebook.com/15704546335/videos/653258435580236	https://m.facebook.com/story.php?story_fbid=653258435580236&id=15704546335
NBC First Debate	https://www.facebook.com/155869377766434/videos/993498417837076	https://m.facebook.com/story.php?story_fbid=993498417837076&id=155869377766434
NBC Second Debate	https://www.facebook.com/155869377766434/videos/261699228607018	https://m.facebook.com/story.php?story_fbid=261699228607018&id=155869377766434
NBC Vice Debate	https://www.facebook.com/155869377766434/videos/342907180316163	https://m.facebook.com/story.php?story_fbid=342907180316163&id=155869377766434

First, we collect the major news outlets that live broadcast debates on Facebook live streams. We find that NBC News, ABC News, and Fox News all live stream 2020 Presidential Debates on their main Facebook page, which covers a wide political spectrum. We gather the URL links to the live streams of these three debates (2 Presidential debates and 1 Vice Presidential debate) for both desktop and mobile versions, which results in the links in Table A1.

Since most comments in live streams only occur during the debate, we are able to collect the comments made during (not after) the debate. To do this, we built a python script using the Selenium Webdriver to click multiple times in the tag "more comments" and expand the comment box. We repeated the same procedure for all the nine links until the memory limits of our webdriver, and then we saved the expanded static Facebook webpage in HTML locally. Finally, we used scrapping techniques through the HTML tags to retrieve the comments and other information displayed in the webpage.

The main limitation of this approach is that our strategy only collects a sample of the available comments. Due to memory limitations of our web drivers, we were not able to expand the comments box and display all the comments made during the specific debate. Therefore, after a few hours, our Selenium Webdriver was not able to load more comments. For this reason, we collected almost the same amount of comments for every debate.

In addition, the use of Selenium was not possible on the desktop URLs; it is only possible through the mobile URL. For this reason, we were able to build a large sample only for the 'unfiltered' comments, since this is the only option on Facebook mobile URLs. In the case of the "most relevant" comments, a filter created by Facebook to facilitate the users experience with streaming chats, we expanded the chatboxes manually, and therefore, ended up with a considerably smaller sample, and use this sample only to assess the robustness of our results.

All the comments collected are public, and we did not retrieve any personal information from the Facebook Users. Therefore, we do not believe our data collection puts any user in risk. And we collected all the samples at the same day, which means that although our sample is not randomly selected of the entire universe of Facebook users, our consistent sampling method of sampling provides reliable information about potential differences within-event between platforms, and within-platform between debates.

Even without any type of randomization in our sampling strategy, we have no reasons to doubt our sample is representative from the overall population of comments made on the live broadcasting events under discussion. First, as our analysis show, our samples are relatively similar to the “most relevant” comments; in terms of toxicity, length, and partisan dynamics, our findings are consistent across these two samples. That being said, we describe our results as speaking for the sample of comments we were able to collect.

Supplemental Results

Figure A1: Number of Characters per Comment

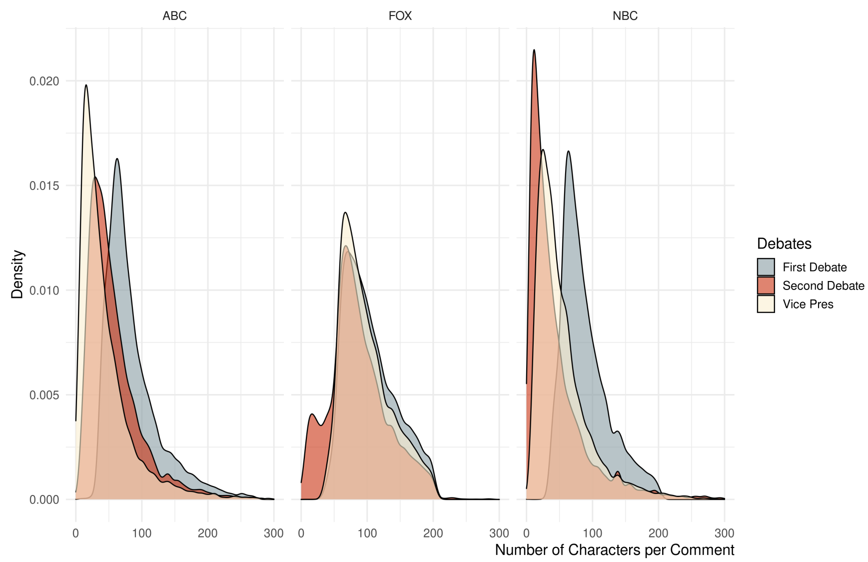


Figure A2: Density of the Toxicity Scores

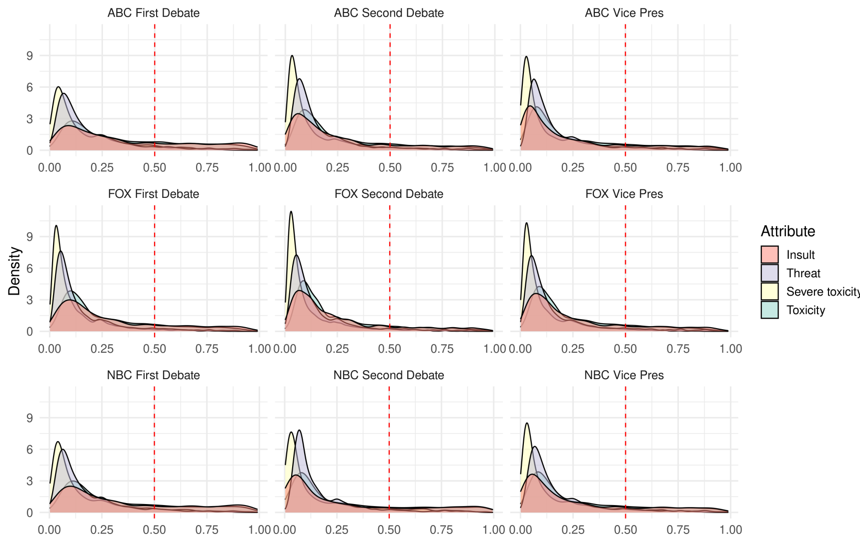


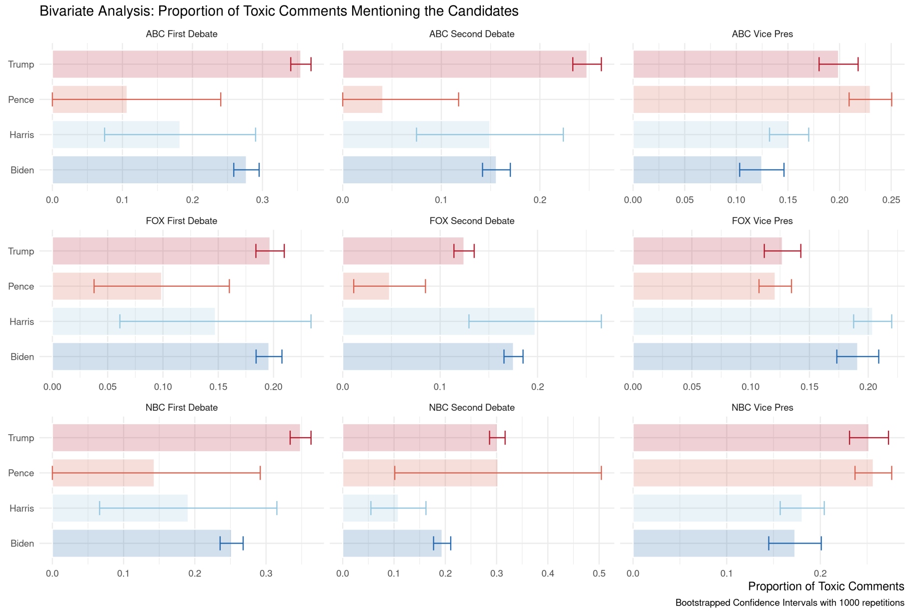
Table A2: Examples of Comments by Toxicity and Insult Level

Toxicity Score Above .8	"Trump is so disrespectful in the debate, He is a pure CLOWN and a BUM!!" "Really!! Lets vote a little girl lover into office!! Joe Biden is a sick human!!"
Toxicity Score Between .5 and .8	"Trump is again acting like a toddler" "Biden is cool and composed compared to juvenile trash-acting trump"
Toxicity Score Below .5	"Mr. President cant you stop talking for a minute and give Mr.Joe Biden a chance to talk?" "Biden is the only choice voting blue"
Insult Score Above .8	"Trump is stupid" "This moron is insane"
Insult Score Between .5 and .8	"Pence is kicking her butt with hard core FACTS!!!" "Biden's words = lame politicians's words"
Insult Score Below .5	"Trump for sure 2020!" "Good Joe go go go"

Table A3: Examples of Comments by Severe Toxicity and Threat Level

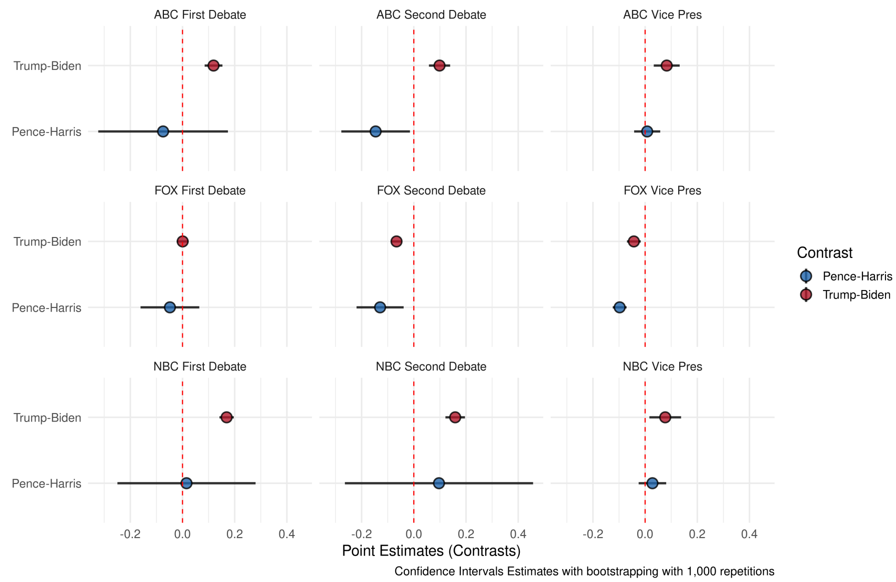
Severe Toxicity Score Above .8	"Mute Trumps fucking mouth" " GREAT JOB VP PENCE!!!!! YOU KILLED IT.... YOU PUT THE DEVIL ON THE RUN"
Severe Toxicity Score Between .5 and .8	"votethatfuckout" "Shut the hell up and quit stopping him talk"
Severe Toxicity Score Below .5	"Trump 2020" "Joe is reading a teleprompter!!"
Threat Score Above .8	"You dems will not be removing our President! We are going to remove you all one by one, the American way!!!!" "Trump supporters shooting up schools in killing people in el paso"
Threat Score Between .5 and .8	"Bernie got fire tonight!" "He won't be re-elected.....he is going to prison..."
Threat Score Below .5	"Show this joke of a president whos boss Joe." "Agreed it is racist. Equality is important not making special cases."

Figure A3: Proportion of Toxic Comments per Candidates, Debate and Platform



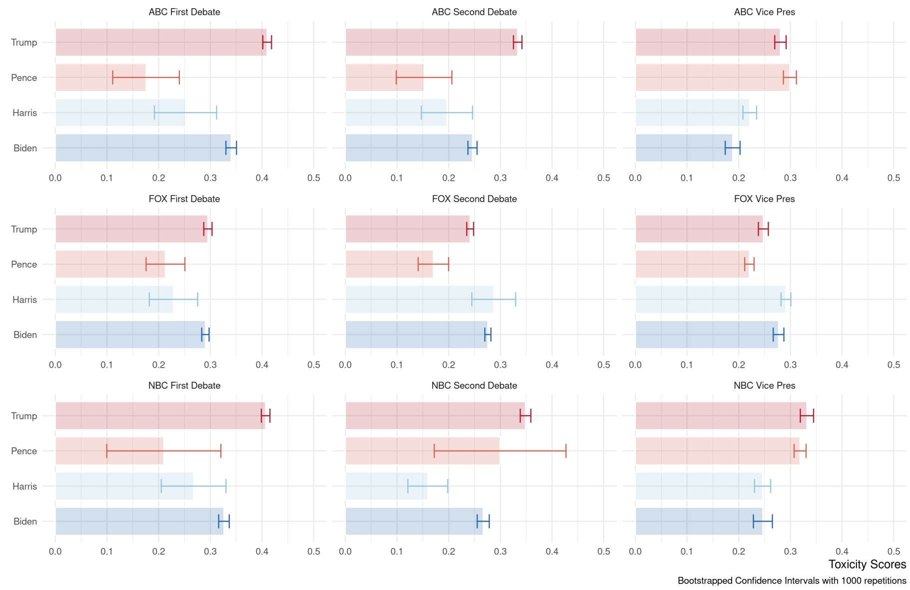
Note: This figure shows the proportion of toxic comments mentioning particular candidates. These analyses do not include any control variables.

Figure A4: Difference in Proportion of Toxic Comments between Candidates, Debate and Platform



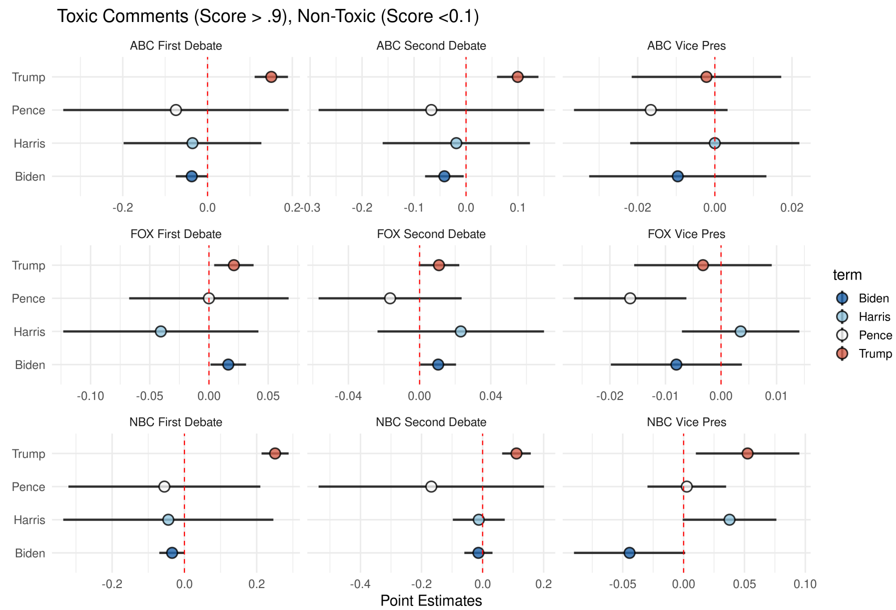
Note: The figure shows the contrasts in toxicity in comments that mention specific political actors. These are contrasts from regression models that also control for the length of comments and number of reactions per comment.

Figure A5: Average Comment Toxicity by Candidate Mentioned, Debate and Platform



Note: This figure shows the average toxicity scores of comments mentioning particular candidates. These analyses do not include any control variables.

Figure A6: Results Under Alternative Cutoff



Note: This figure displays the point estimates of the likelihood a comment is toxic if it mentions vs. does not mention a particular candidate. These come from regression models that control for the length of comments and number of reactions a particular comment receives. The results generally reflect the same pattern as those using the .5 threshold, but the confidence intervals are much larger given the smaller number of comments that are above .9 or below .1.

Figure A7: Robustness Checks: Results Using Insults, Threats and Severe Toxicity

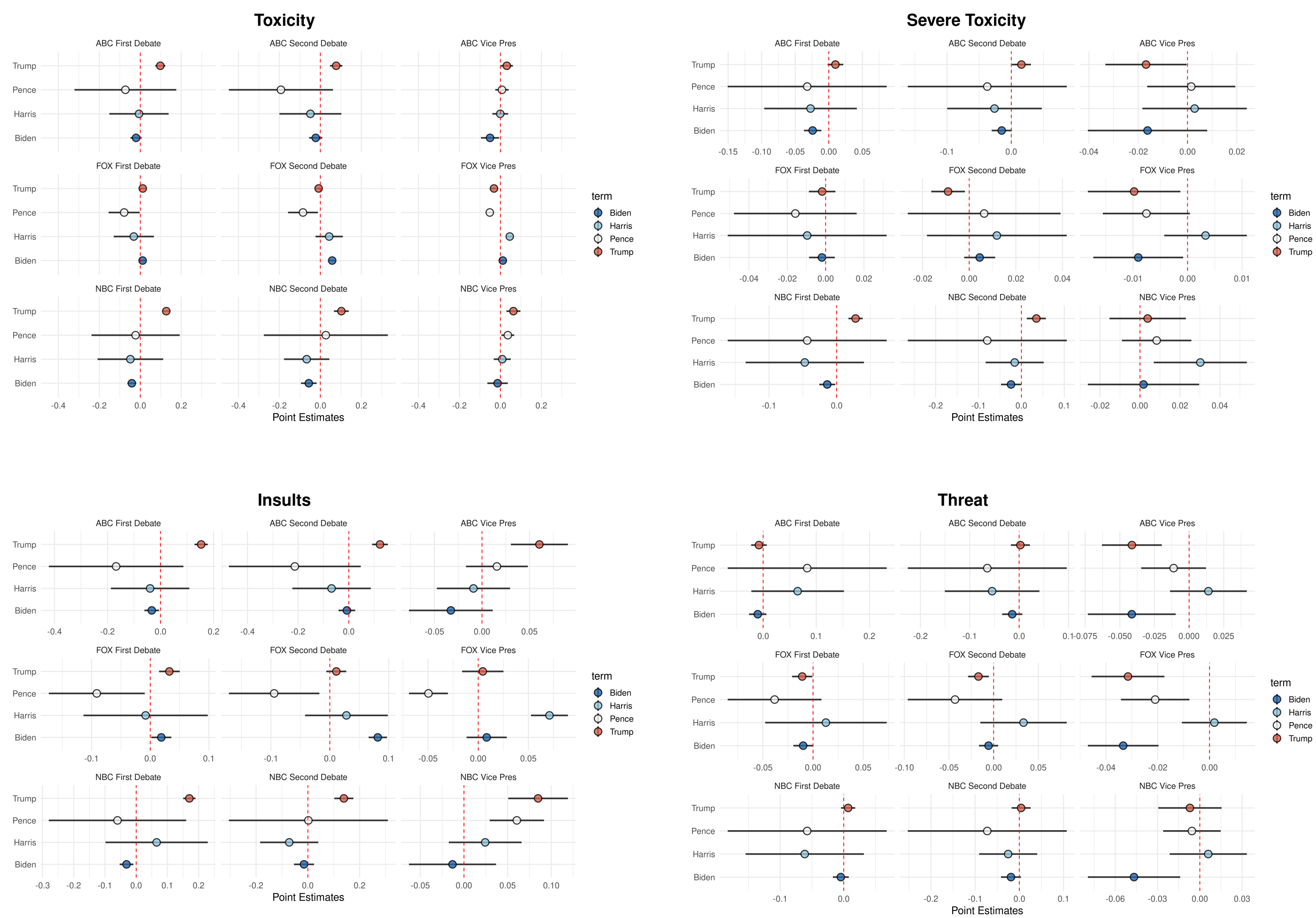


Table A4: Regression Models:

	Dependent Variable: Toxicity Scores								
	ABC			FOX			NBC		
	First Debate	Second	Vice-Pres	First Debate	Second	Vice-Pres	First Debate	Second	Vice-Pres
Intercept	0.156^{***}	0.137^{***}	0.097^{***}	0.092^{***}	0.031^{***}	0.079^{***}	0.153^{***}	0.137^{***}	0.104^{***}
	(0.017)	(0.019)	(0.015)	(0.012)	(0.011)	(0.011)	(0.014)	(0.017)	(0.016)
Biden (mentions)	−0.021	−0.023	-0.051^{**}	0.011	0.058^{***}	0.012	-0.041^{***}	-0.056^{***}	−0.014
	(0.014)	(0.016)	(0.023)	(0.008)	(0.008)	(0.010)	(0.011)	(0.019)	(0.025)
Harris (mentions)	−0.006	−0.049	−0.001	−0.032	0.043	0.046^{***}	−0.048	−0.067	0.009
	(0.074)	(0.077)	(0.020)	(0.050)	(0.034)	(0.009)	(0.082)	(0.056)	(0.021)
Trump (mentions)	0.098^{***}	0.077^{***}	0.031^{**}	0.012	−0.008	-0.031^{***}	0.127^{***}	0.103^{***}	0.063^{***}
	(0.012)	(0.015)	(0.015)	(0.008)	(0.008)	(0.010)	(0.010)	(0.019)	(0.017)
Pence (mentions)	−0.073	−0.192	0.008	-0.079^{**}	-0.085^{**}	-0.052^{***}	−0.023	0.027	0.036^{**}
	(0.127)	(0.130)	(0.017)	(0.039)	(0.037)	(0.009)	(0.110)	(0.154)	(0.016)
User's Interactions	0.004^{***}	0.003^{**}	−0.001	0.0004^{**}	0.0003	0.001^{***}	0.0005^{**}	0.002^{*}	0.002
	(0.001)	(0.001)	(0.002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.001)	(0.001)
Number of Words	0.003^{***}	0.001	0.007^{***}	0.004^{***}	0.004^{***}	0.004^{***}	0.004^{***}	0.005^{***}	0.006^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	5,016	2,832	3,147	9,329	8,096	9,187	8,060	2,319	3,281
Adjusted R^2	0.021	0.014	0.027	0.007	0.016	0.016	0.027	0.050	0.018

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table A5: Regression Models: Bivariate Linear Probability Models (No Controls)

	Dependent Variable: Toxicity Scores								
	ABC			FOX			NBC		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.235***	0.160***	0.192***	0.165***	0.113***	0.161***	0.213***	0.216***	0.191***
	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.008)	(0.005)
Biden (mentions)	−0.005	−0.020**	−0.057***	0.024***	0.064***	0.019**	−0.015	−0.031***	−0.023
	(0.010)	(0.009)	(0.015)	(0.008)	(0.006)	(0.010)	(0.010)	(0.012)	(0.016)
Harris (mentions)	−0.070	−0.005	−0.023*	−0.042	0.035	0.053***	−0.027	−0.092**	−0.017
	(0.064)	(0.042)	(0.013)	(0.049)	(0.030)	(0.009)	(0.073)	(0.039)	(0.013)
Trump (mentions)	0.121***	0.091***	0.003	0.022***	−0.010	−0.029***	0.139***	0.089***	0.046***
	(0.009)	(0.008)	(0.010)	(0.008)	(0.007)	(0.010)	(0.009)	(0.010)	(0.011)
Pence (mentions)	−0.171*	−0.167**	0.041***	−0.084**	−0.079**	−0.048***	−0.068	0.068	0.057***
	(0.103)	(0.077)	(0.011)	(0.039)	(0.031)	(0.009)	(0.097)	(0.095)	(0.010)
N	9,855	9,955	9,448	9,786	12,463	9,824	9,905	7,661	9,922
Adjusted R^2	0.018	0.014	0.005	0.002	0.010	0.009	0.024	0.015	0.006

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Results for Most Relevant Comments

Table A6: Most Relevant Comment Sample

Facebook Channel	# Comments	# Average Length of Comments
ABC First Debate	539	19.01
ABC Second Debate	488	17.31
ABC Vice Debate	557	16.41
FOX First Debate	619	17.27
FOX Second Debate	604	17.38
FOX Vice Debate	613	16.69
NBC First Debate	485	18.21
NBC Second Debate	466	16.47
NBC Vice Debate	348	17.76

Figure A8: Most Relevant: Proportion of Comments Mentioning a Candidate

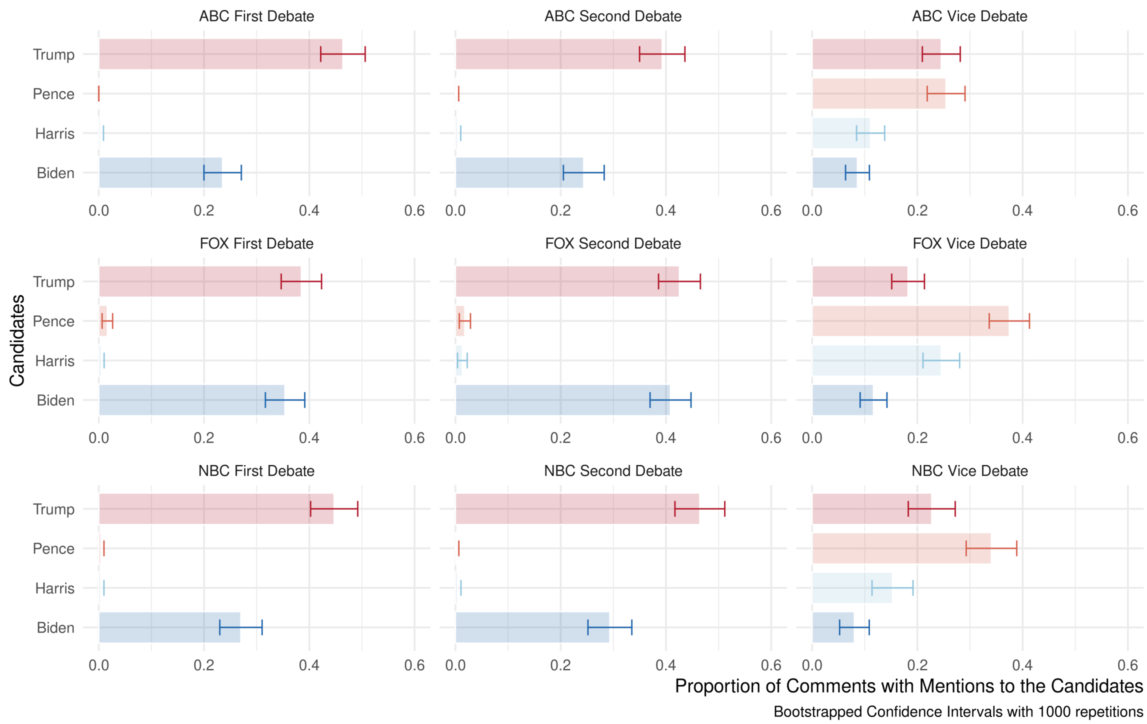


Figure A9: Most Relevant: Proportion of Comments that are Toxic or Insulting by Debate and Platform

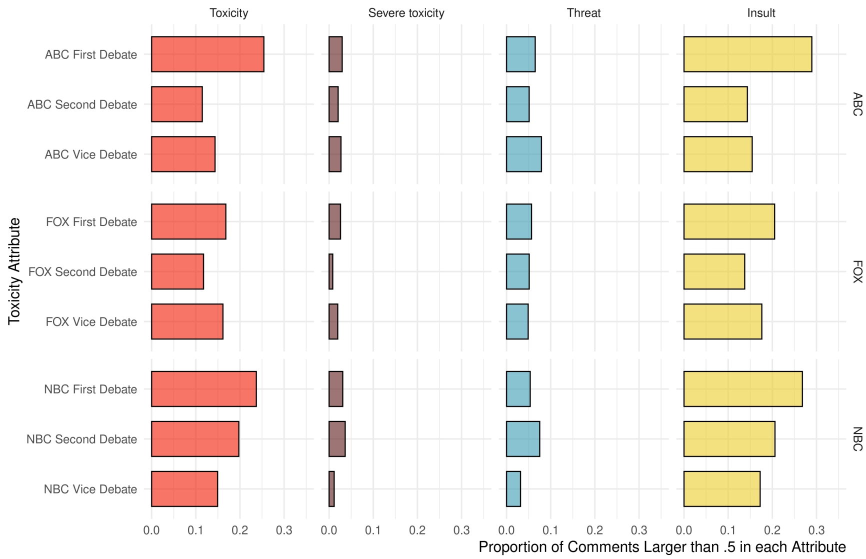
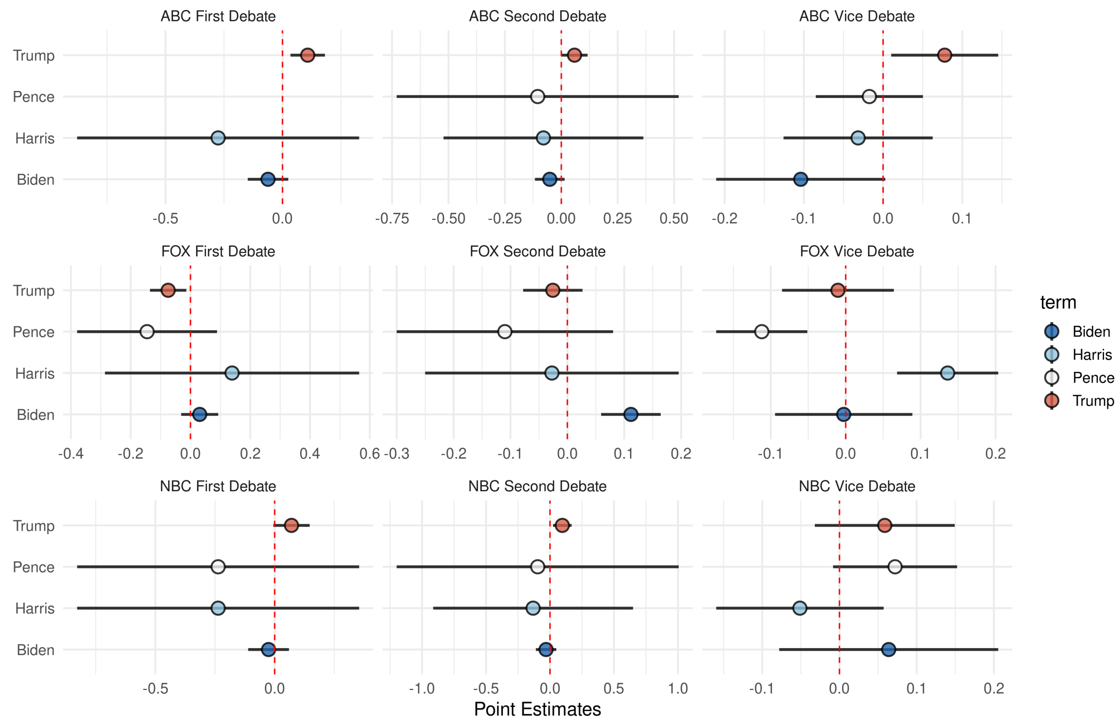


Figure A10: Most Relevant: Toxicity by Candidate, Debate and Platform

Figure A11: Most Relevant: Predicting Toxic Comments, by Debate and Platform



Censoring of Toxic Comments

It is possible that the media outlets activated moderation tools to facilitate the censoring of some extremely toxic comments as the live broadcast is happening. We take two strategies to explore evidence of moderation: 1) Looking for discontinuities in the distribution of toxicity scores and 2) Using a dictionary analysis to quantify the amount of profanity in the chat streams.

Our first strategy to find evidence of censoring on streaming chats combines the values retrieved from the toxicity scores and a statistical procedure to detect change points in the density of the comments. First, we bin the scores for every 0.05 interval between 0 and 1, then, we sum the number of comments for every bin. Using this binned data, we estimate discontinuity models using a local linear regression approach and robust confidence intervals, as in (Titiunik et al., 2015), using every 0.10 interval between 0.1 and 0.9 in the toxicity scores as the cutoff placebos. Our hypothesis is that, in the case of censoring, we would find negative point estimates - a perceptive decrease in the number of comments - at the right extreme of the distribution, when comments are more toxic.

The results for this validation check are presented below. We find no evidence of censoring on the right extreme of the toxicity distribution. In other words, our analysis finds no abrupt decrease in the number of comments as one moves along toward extremely toxic scale. We perform the analysis splitting the data by platform.

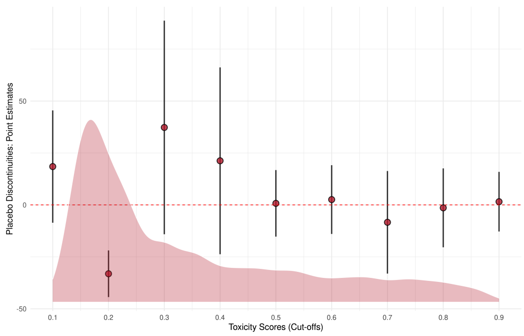


Figure A12: ABC Debates

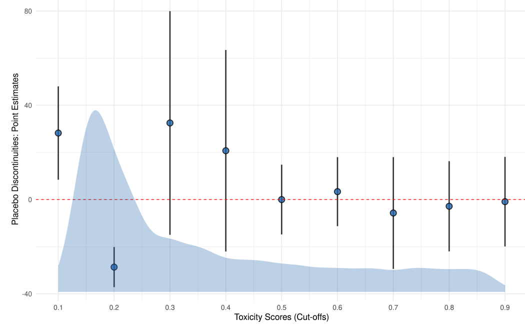


Figure A13: NBC Debates

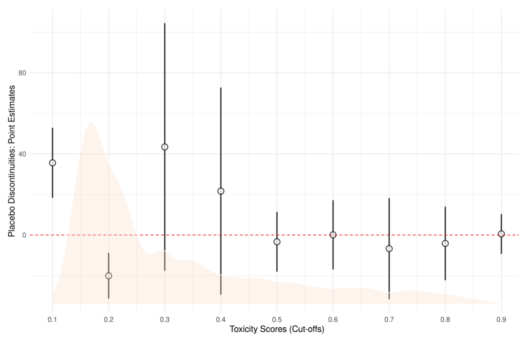
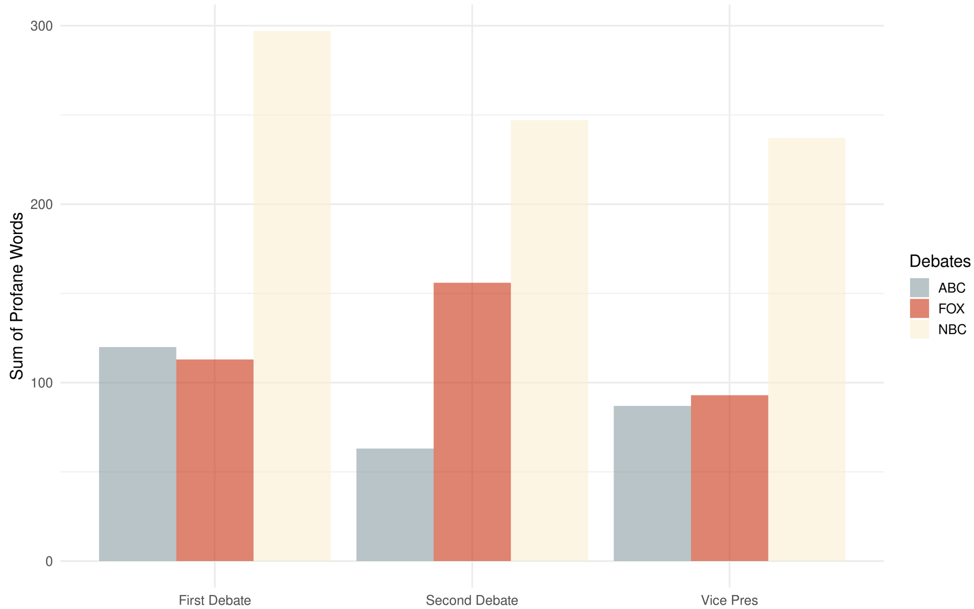


Figure A14: Fox Debates

Note: The plots show placebo tests to detect abrupt discontinuities in the number of comments across the toxicity scores. We use every .1 interval between 0.05 and 0.95 for the toxicity scores as cut-off placebos for the regression discontinuity estimation. Point Estimates and Confidence Intervals estimates using a local linear regressions, robust standard errors and data-driven bandwidth selection (Titiunik et al., 2015). The density for the scores are presented in the background for each graph.

There are many ways for comments to be toxic. Therefore, the presence of high levels of toxicity does not rule out the possibility that other filters were present. As a second empirical analysis of censoring, we apply a dictionary analysis to the comments in each debate stream using the lexicon of profane words in the "sentimentr" R package (Rinker, 2019). Figure A15 shows the number of profane words found in each stream (unweighted). Overall, the use of profanity is relatively low compared to the total number of words in a given stream's sample, which range from approximately 70,000 to more than 200,000.

Figure A15: Profanity Words in the Streaming Chats



NBC streams have markedly higher levels of profane words than the ABC and Fox streams, suggesting that there may be a difference in the moderation. We would not necessarily expect there to be zero uses of profane words present in the streams because the filters used by Facebook likely do not match the dictionary employed in this analysis exactly. As a concrete example, we find instances of the word "f***ing" in the NBC stream, while only misspellings or alternative uses of the word (e.g., "f*cc", "clusterf**k") in the ABC stream or Fox stream (e.g., "shutthef***up"). This highlights the blunt nature of these types of filters, which still permit the presence of toxic comments, such as "really! like american's are idiots !! we don't need kamala telling us what the word debt means.. shutthef***up liar!" which was allowed on the Fox stream and scored as a toxic comment in our empirical analysis.

In addition to limiting profanity, it is possible that the moderators limited the use of other specific phrases from their streams. This is more difficult to detect. However, we can explore the use of common phrases that are in the political discourse to get an initial sense if it appears these types of tools were present. For example, we searched for the use of "fake news" in the comments. While this phrase was present anywhere between 4 and 31 times in the ABC and Fox streams, it had zero mentions in the NBC streams. In contrast, the popular Trump slogan "MAGA" was present across channels, as was the criticism, "sleepy Joe." This suggests there may have been some use of selective moderation, but it was not absolute and did not prevent commenters from posting otherwise toxic commentary or many of the popular phrases in the political discourse.