Journal of Communication ISSN 0021-9916

ORIGINAL ARTICLE

The Effect of Streaming Chat on Perceptions of Political Debates

Victoria Asbury-Kimmel[1], Keng-Chi Chang[2], Katherine T. McCabe[3], Kevin Munger[4], and Tiago Ventura[5]

[1]Department of Sociology, Harvard University, Cambridge, MA 02138, USA

[2]Department of Political Science, University of California San Diego, La Jolla, CA 92093, USA

[3]Department of Political Science, Rutgers University, New Brunswick, NJ 08901, USA

[4]Department of Political Science and Social Data Analytics, Pennsylvania State University, State College, PA 19165, USA

[5]Department of Government and Politics, University of Maryland, College Park, MD 20742, USA

Broadcast media consumption is becoming more social. Many online video "livestreams" come with embedded livestreaming chatboxes, uniting the on-screen and social components. We investigate how streaming chat shapes perceptions of political events. We conducted a field experiment during the September 2019 Democratic Primary Debate where subjects were assigned to view the debate with or without streaming chat. We use text analyses to characterize the frequency, toxicity, and tone of comments in the chat. Our experimental findings indicate that Democratic subjects assigned to the Facebook (social) chat condition reported lower affect toward Democrats and a worse viewing experience, aligned with the toxic and overwhelming nature of the chat. The polarity of candidate-directed comments also influenced candidate evaluations and perceived performance in the polls. This suggests that consumers of mass media will be both more immediately affected by social feedback and likely to make inferences about the experiences of their fellow consumers.

Keywords: Social Media; Second Screening; Experiment; Debates; Media Multitasking; Text Analysis

doi: 10.1093/joc/jqab041

Media consumption is becoming more social. Media producers encourage audiences to engage with their content via social media, and media consumers take to social media as "second screens" to see what others think as they watch various types of content, ranging from pre-recorded season finales to live events. More

*Corresponding author: Katherine T. McCabe; e-mail: k.mccabe@rutgers.edu

recently, the advent of live video with integrated streaming chat is gaining in popularity among 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.

In the political arena, young politicians like Congresswoman Alexandria Ocasio-Cortez have famously adopted livestreams with streaming chat as a strategic tool, and major political events are increasingly being broadcast through platforms that offer streaming chat options. Facebook reported that ABC News received 6.2 million views while livestreaming the first political debate between Donald Trump and Hillary Clinton in 2016, and Fox News received nearly 370,000 comments during their livestream of the debate on the platform.[1] By the 2019–2020 election season, multiple Democratic primary and all general U.S. presidential election debates were livestreamed on at least one platform that provided streaming chat.

This rapid technological change has the potential to modify the effect of live broadcasts on the viewing public by altering aspects of the experience that had been constant for decades. Previous scholars have shown that exposure to commentary about political events can alter perceptions of these events. The rise of "dual" or "second screening"—where viewers follow along with social commentary on a second device during the live broadcast—offers an enhanced source of influence in real time (Gil de Zúñiga et al., 2015; Vaccari et al., 2015). However, despite the recent rise in integrated streaming chat, we have little evidence of the effects of this technology on viewers' experiences and subsequent attitudes or beliefs.

We employ the "Mix of Attributes" approach (Eveland, 2003) for describing new media technology and theorizing about its effects. We identify four salient attributes on which streaming chat is distinct from other media technologies: *frequency* of the comments, *toxicity* of those comments, *content* of comments, and the *context* in which comments were created and consumed. We then provide a digital field experiment to test streaming chat's effects on attitudes.

During the September 2019 Democratic Primary Debate, we pre-registered[2] and conducted a digital field experiment, modeled after the Gross et al. (2019) design, to study the influence of streaming chat on the public's perceptions of the debate and the participating candidates. In a two-wave survey, using Amazon's Mechanical Turk (MTurk) to recruit subjects, we randomly assigned and encouraged 1,095 participants to watch the debate on one of three online platforms: the ABC News website, which provided a livestream of the debate without a streaming chat; the ABC News Facebook page, which provided social commentary from Facebook users alongside the video; and the FiveThirtyEight website, which provided live expert commentary from political analysts alongside the video. We collected the comment feeds from the social and expert conditions to characterize how the chats differed between conditions.

The experimental design aims for high ecological validity: Subjects consume a prominent political event in real time, from the comfort of their own homes; we

merely prompt them to vary the platform with which they view the debate. Our sample consists of people who self-reported they were both planning to watch the debate and had the capacity to watch the debate online. This is not, of course, representative of the general population, but is epistemically desirable. Forcing media upon people who would never choose to consume it produces potentially misleading counterfactuals (Arceneaux & Johnson, 2013).

Our text analysis results reveal that the "social" commentary on the Facebook chat contained a substantial number of negative primes about each candidate and high levels of toxic language. The results of our randomized experiment tightly match these descriptive text analysis results, pointing to the content of comments as one important potential mechanism of the treatment. Democratic respondents who were encouraged to watch the debate on the ABC News Facebook page came away from the debate with more negative feelings toward Democrats, aligning with the negative depictions of the debate participants in the comments, relative to respondents encouraged to watch the debate without streaming chat. The number of negative comments about a given candidate is strongly predictive of decreased feeling thermometer evaluations of that candidate. On the other hand, the number of positive comments about a candidate was highly correlated with the perception that a candidate would do better in the polls post-debate. The "expert" chat had markedly few negative primes or toxic comments, limiting its influence through these channels.

The development of new technologies for people to interact with others while consuming media is undoubtedly exciting in its potential to stimulate viewers. We anticipated the potential for positive effects of the social chat on engagement. However, our findings point out potential pitfalls of real-time comments in disrupting learning and information processing as citizens are exposed to the carnivalesque thoughts of their peers online.

Introduction of streaming chat

The technological capacity for audiences to engage in mass communication on one screen while consuming broadcast news media or a "media event" on another has existed for well over a decade. Though earlier studies documented the existence and prevalence of the phenomenon (Jungherr, 2014; Larsson & Moe, 2012), theoretical interest in dual or double screening began in earnest with Gil de Zúñiga et al. (2015) and Vaccari et al. (2015). "Dual screening" or "second screening," where viewers of political television media participate in real-time conversations on online platforms like Twitter has been shown to change the effect of media consumption on attitudes, political knowledge, and political expression (Barnidge et al., 2017; McGregor & Mourão, 2017; Ran & Yamamoto, 2019; Shah et al., 2016; Vaccari et al., 2015). Traditionally in political communication, the news media are thought to have the power to set the agenda and change the criteria by which consumers evaluate political figures and form attitudes (Iyengar & Kinder, 1987). The addition of a second

screen could, at minimum, muddle the effects of media on public opinion (Camaj & Northup, 2019). However, the extent to which this occurs depends on the amount of attention consumers give to the "second screen," among other factors.

Our study focuses on an innovation to dual-screening: streaming chatboxes alongside a related video feed, all on one screen. We argue that "integrated real-time streaming chat" (hereafter, "streaming chat") may amplify the effects of the second screen by immersing viewers in both the video livestream and the real-time comments simultaneously—shaping what is salient and accessible as consumers form their attitudes.

Inspired by the "Mix of Attributes" approach to the study of new communication technologies advocated by Eveland (2003), we propose an original theoretical framework of four pathways by which the addition of streaming chat to broadcast media might affect perceptions of the media event: the frequency, toxicity, content, and context of the comments. Our discussion is premised on a technosocial context similar to the Facebook "social chat" in our study, but we hope that identifying theoretically distinct pathways will extend the temporal validity of our study in what is a rapidly evolving communication technology (Munger, 2019).

- Frequency: High volume of streaming comments increases distraction and information overload.
- Toxicity: Extreme negative affect changes the overall emotional experience.
- Content: Discussion topics serve as primes, increasing their salience.
- Context: The composition of commenters is not obvious to the viewer, leading them to make misleading inferences about public opinion.

We now trace the theoretical antecedents of each of these pathways and hypothesize their potential consequences within the context of the setting of our study: the September 2019 U.S. Democratic Primary Debate. As we explain in detail below, the treatment conditions assign subjects to view two different types of streaming chat—on Facebook (Social condition) or on FiveThirtyEight (Expert condition)—which vary each of these parameters. These conditions are compared to the control: viewing on a platform that does *not* have a streaming chat. "Expert chat" is a much less prominent phenomenon than social chat, so while we consider social chat our primary condition of interest, the expert chat serves as a useful additional benchmark.[3]

Our hypotheses and analyses focus on particular theoretical dimensions associated with the experience of viewing the debate with streaming chat on Facebook, with expert chat, or without streaming commentary. Given the ecological validity of these natural settings, it is impossible to enumerate all the dimensions on which the field experimental conditions differ. Any viewing experience comes with a bundle of attributes that are difficult to fully disentangle. In addition to the theoretical attributes of interest, there are three other attributes that bear mention. While we do not have the empirical leverage to test the effects of these attributes here, they serve both

as scope conditions, elaborating on the specific nature of the conditions in our study, as well as potential areas for future research.

- Interactivity: The capacity for viewers to actively participate in the streaming chat by posting comments or "Liking" other comments.
- Textuality: The form the comments can take, including the possibility of embedding emojis, images, or gifs.
- Broadcast visibility: The relative distribution of the different elements, primarily the ratio of screenshare given to the broadcast compared to the streaming chat.

Given the capacity to *interact* with the content, we would expect the social chat to increase engagement and affective response, but we lack empirical evidence about which subjects actively commented or "liked" other comments in the chat. Textuality refers to the range of encoding options for posted commentary. Does the platform allow for emoji, embedded gifs, or "reactions" (e.g., Facebook "Angry" or "Love" buttons)? These attributes supplement text-like communication with affective, relational information. This was possible in the social chat but not the expert chat, although the latter did enable commenters to include statistical figures. Our theoretical expectation here is that *textuality* allows for an increased range and intensity of affective communication. We did not collect non-text data, so we lack empirical leverage on this dimension. Finally, *broadcast visibility* is likely to be important, given the empirical range of this attribute across the three conditions. In both the control and social chats, the broadcast typically takes up roughly two-thirds of the screen. However, in the expert chat, the debate video is constrained to a sidebar, no more than one-sixth of the screen. Some subjects complained that this made it difficult to see the debate. Ultimately, we do not believe there is theoretical insight here beyond the fact that people do not like to squint, but this attribute is also likely to affect enjoyment.

Frequency: Distraction

Some early research on second screening focused on the capacity of the second screen to distract the audience from the media event (Gottfried et al., 2017; Van Cauwenberge et al., 2014) by providing too many stimuli for viewers to direct their attention. However, research that centers the purposiveness of second screening also emphasizes the agency of viewers in selecting when and where to seek out the second screen (McGregor & Mourão, 2017), and whether to participate (Vaccari et al., 2015), reducing the potential independent effects of second screening.

In contrast, the unity of the broadcast and the chat in the visual field makes streaming chat theoretically distinct from second screening, and potentially more likely to affect viewers and the viewing experience. When an individual watches a live video with streaming chat, the chat is often located at the same eye level as the

video broadcast, allowing viewers to consume both automatically. This increase in stimuli may serve to increase *engagement* with the video content. For example, Facebook reports that people comment 10 times more on their platform's live videos than non-live video content (Greenberg, 2016). Moreover, an alternative strand of research concerning a more general form of dual screening—media multitasking—provides a different perspective that supports the possibility that streaming commentary may have benefits even if it reduces cognitive performance. Wang and Tchernev (2012) suggest that people consume multiple streams of media at once when they are nominally engaged in information seeking, because multitasking provides unanticipated emotional gratifications.

However, because integrated streaming chat also undercuts the purposiveness that even moving the eye between two screens requires, it may simultaneously increase the relevance of concerns about distraction. Even if viewers intend to focus on the video, viewers may be incidentally exposed to the real-time comments, even if they did not intend to seek out comments. Concerns about the potential of distraction from streaming chat are particularly important in the context of this study because one of the primary functions of presidential primary debates is to provide candidate information (Benoit et al., 2002). On the one hand, exposure to streaming chats during debates could serve a normatively desirable purpose as an additional source of information. However, the frequency of comments in the stream might also undermine what a viewer gets out of the experience.[4] The viewer may find watching a broadcast with streaming chat more distracting and demanding, and therefore, less *informative* or *enjoyable*.

Although "enjoyment" is an intrinsically subjective phenomenon, there are a number of ways to conceptualize "informativeness" and "engagement." For instance, engagement can be defined in terms of observable behaviors such as "liking" comments, "tweeting along" (Jennings et al., 2020a), or asking questions of the candidates (McKinney & Rill, 2009); it can also be defined as one's personal assessment of how much attention was given to the debate itself (Houston et al., 2013). Similarly, "informativeness" can be defined in terms of informational recall (Holbrook, 1999; Jennings et al., 2020b; Weaver & Drew, 1995) or judgments about (perceptions of) the candidates and their positions (Warner et al., 2011; Zhu et al., 1994). In the present study, we focus on respondents' self-reported assessments of the "enjoyment" they experienced from watching the debate, perceived "informativeness" of the debate, and their subjective sense of "engagement" with the debate, similar to the conceptual definition in Houston et al. (2013).

We focus on perception because, in the realm of politics, perception is power. Primary debates, in particular, have the capacity to alter viewers' perceptions, which can influence downstream candidate strategies and voting intentions of the electorate (Benoit et al., 2003; Shockley-Zalabak et al., 2019). Specifically, the meta-study of Benoit et al. (2003) suggests that general debates mainly affect perceptions of issue salience (but not candidate competence or personal character), while primary debates can affect perceptions of candidate competence and character. Similarly,

Shockley-Zalabak et al. (2019) focuses on trust perceptions toward candidates in different issue areas. The authors suggest that Trump was more trusted regarding issues salient in swing states, influencing voter decisions and contributing to his Electoral College vote.

In summary, we hypothesize that the *frequency* element of the streaming chat platform design will change how viewers experience the media event relative to contexts where viewers can focus exclusively on the video without streaming commentary:

> Hypothesis 1: Increased distraction from the streaming chat will cause viewers to consider the debate less enjoyable and informative, but the increased number of stimuli will cause them to be more engaged, relative to watching without streaming chat.

Expert chats typically have a very low frequency of comments, decreasing the problem of distraction. The addition of a small number of comments, we predict, will only enhance how enjoyable and informative viewers perceive the experience to be, while also adding stimuli that can increase engagement relative to viewing the debate without any streaming commentary. Research on media multitasking supports this possibility: Shin et al. (2020) find that moderate amounts of media multitasking—degrees of task switching that are within the range of our expert chat condition—may actually enhance attention.

> Hypothesis 1(e): Commentary from the expert chat will cause viewers to consider the debate more enjoyable, informative, and engaging (relative to no streaming chat).

Toxicity: Affective processing

Our next set of hypotheses stems from the premise that streaming social chats are more likely to be composed of low-quality, toxic comments than other forms of media due to the anonymity or quasi-anonymity of comments. The stream in the Facebook chat is made quasi-anonymous by the speed at which it flows. In the first block of the debate, for example, we estimate there were 60.3 comments per minute. While viewers might notice frequent commenters, there are too many comments for long-term reputational costs. This quasi-anonymity has implications for the types of people who comprise the streaming chat and for the types of messages they send. For instance, anonymity increases "flaming" or personal attacks in online communities (Mungeam, 2011), and removing anonymity elevates civility in online newspaper comments sections (Santana, 2014). The presence of toxicity in a chat can produce a feedback loop, as people with a low tolerance for uncivil or impolite language opt-out of sending messages (Theocharis et al., 2016), and joining an online community where toxic language is normalized causes people to comment in a

more toxic manner (Cheng et al., 2017; Rajadesingan et al., 2020). As we demonstrate below, streaming Facebook chats can be highly toxic.

We define "toxic" comments as rude or disrespectful remarks and attacks. Toxicity is closely related to and may be conceptualized as a component of "incivility," which has been studied widely and as Coe et al. (2014) note, is a "notoriously difficult term to define" (p. 660). There is a broad literature distinguishing different types of incivility (e.g., Chen, 2017; Gervais, 2015; Masullo Chen et al., 2019; Papacharissi, 2004). Our theoretical interest is in the emotional reactions to toxic commentary, which we believe is based primarily on whether a participant perceives a comment to be rude or disrespectful, and less so driven by the precise content of the toxic remark. Our concept and definition of toxicity, as perceived rude or disrespectful remarks and attacks, follow conceptual definitions in research by Coe et al. (2014) and Kim and Kim (2019), as well as Kim et al. (2021) who use the same empirical measure of toxicity that we will use in this study in their study of social media comments.

Recent research suggests that political media consumption in the contemporary climate of total saturation and extreme affect can cause people to feel overwhelmed, anxious, and angry (Wagner & Boczkowski, 2019). Furthermore, Kim and Kim (2019) and Gervais (2015) also find that exposure to incivility in political discussion leads to these negative emotions. Kim and Kim's (2019) conceptualization of incivility—"disrespectful statements for the purpose of attacking" (p. 220), closely aligns with our definition of toxicity, as does the Gervais (2015) definition of uncivil political talk, which includes "the presence of disrespect, hyperbole, and histrionic presentations" (p. 3). Given the wealth of evidence for the central role these emotions play in political communication and participation (see Wagner and Morisi, 2019 for a summary), we pre-registered a hypothesis to examine anger and anxiety as outcomes:

> Hypothesis 2: Streaming chat will cause viewers to be angrier and more anxious relative to viewing experiences without streaming chat.

In contrast, we anticipated that there would be very low levels of toxicity in the expert chat condition; this was borne out in our empirical analysis below. We thus expected that the expert chat would be less likely to trigger negative emotional reactions. Instead, given that expert comments may lead individuals to be more reflective during the debate, our hypothesis is:

> Hypothesis 2(e): The expert chat will cause viewers to be less angry and anxious relative to viewing experiences without streaming chat.

Content: Priming

One of the processes most central to contemporary theories of political communication is priming. Priming changes the relative salience of different considerations

determining how viewers evaluate politicians, where considerations refer to any criteria that might lead someone to favor or oppose a candidate, including issues or personal characteristics. Theoretically grounded in the cognitive process of "spreading activation," priming affects which thoughts are immediately salient when evaluating a given politician (Iyengar, 1987).

The concept of priming in media effects was developed for the broadcast era. Within the debate setting, scholarship on priming effects was primarily focused on commentary immediately before or after a debate broadcast. Newspaper journalists were initially the chief political actors that signaled what information in a debate mattered. As cable news became more relevant, cable news hosts and commentators became important communicators of how debate performances should be interpreted (Fridkin et al., 2008). The rise of dual-screening and then streaming chat has the potential to break down journalistic gatekeeping and influence perceptions in real-time. Second screening and streaming chat enable near-instantaneous priming, as the viewer can consume commentary on a given debate response while it is ongoing. Past work has shown that second screening increases the likelihood of persuasion among consumers of broadcast media (Barnidge et al., 2017), likely due to the immediacy of audience feedback.

Furthermore, the primes produced in unmoderated social chat potentially include primes that would not be produced by traditional media sources. This includes primes that are openly racist, sexist, or otherwise discriminatory, and primes that are untrue or, bluntly, asinine. Holding the content of a news story's summary constant, the addition of social commentary that misrepresents that summary can cause people to recall the incorrect information rather than the story content (Anspach & Carlson, 2020). Social information is more salient and memorable.

In many cases, the primes in the debate comments in our study are bad for the democratic process, encouraging voters to evaluate candidates on criteria that are unrelated to their performance and reinforcing existing inequalities. Consider the example of Biden's acuity. During the debate in this study, Julian Castro specifically criticized Biden's age, and many comments in the streaming chat mentioned Biden's health or acuity at various points when Biden was talking. This immediacy could supplant whatever policy position Biden was attempting to link himself to, instead priming concern about his fitness for office.

On this pathway, the streaming social and expert chats are substantively similar in terms of theoretical predictions. We anticipate that the frequency and valence of candidate-specific primes will influence candidate evaluations. We anticipate that prominent negative primes about the candidates will lead to a decrease in evaluations of the candidates who are the focus of these negative comments. This hypothesis was not pre-registered and could not be fully specified in advance, as the candidate-specific predictions depended on the results of the comment analysis.

> Hypothesis 3: Viewers of the streaming chat will reduce their evaluation of candidates in the presence of a prominent negative prime.

Hypothesis 3(e): Viewers of the expert chat will reduce their evaluation of candidates in the presence of a prominent negative prime.

While discussed as "Content" effects, the *frequency* of comments in the social chat could also enhance the likelihood of the negative priming effects for the social streaming chat, in particular. The frequency of social commentary makes reasoning or counter-argument against these negative messages prohibitive, potentially making viewers more susceptible to influence.

Context: Inferences about the public

Barnidge et al. (2017) demonstrate the importance of second screening by showing how it enhances persuasion among viewers of political content. This occurs because the second screen provides "social cues . . . [which] provide direct evidence of social opinion, which people take to be representative of public opinion, even if they are not" (p. 313). This last claim references Lerman et al. (2016), who demonstrate the existence of the "majority illusion" that can be produced by social networks. Social media contexts are much "denser" than traditional media or socialization; the consumer can observe a large number of signals in a very short period of time. The issue is that our intuitions about the informational content of each signal are misleading because the signals are highly correlated. This tends to magnify belief updating, resulting in what Ortoleva and Snowberg (2015) describe as overconfidence, which causes ideological extremeness.

In the context of this study, subjects exposed to streaming social chat are likely exposed to precisely those conditions that could cause misleading inferences about public opinion: a high frequency of quasi-anonymous comments. The large number of signals entices the viewer to make an inference about public opinion, but it is far too cognitively taxing to do so "correctly"—to accurately account for the fact that signals are being sent by a small and non-representative group of commenters. In the context of a political debate, viewers may come away with markedly different impressions of candidate performances based on the nature of comments each candidate receives. This could influence perceptions of how much the debate improves or hurts a candidate's viability—perceptions of how well a candidate will do in the polls after the debate. For primary election campaigns, in particular, viability is a key factor that influences vote preference (Yawn et al., 1998). While the debate itself can influence perceptions of candidate performance and viability (Yawn et al., 1998), we anticipate that streaming commentary will provide additional (possibly misleading) signals to the consumer about how the broader public perceives candidate performance.

The large number of quasi-anonymous comments in *social* streaming commentary is expected to influence perceptions of broader public opinion, but, in contrast, the expert chat clearly represents itself as such. We, therefore, expect that the expert

chat should not cause viewers to change their opinion of the *mass public's* opinions about candidate performance and viability.[5] This produces the hypotheses:

> Hypothesis 4: Respondents will change their estimate of the success of each candidate in future polls based on the sentiment directed at that candidate in the streaming chat.
> Hypothesis 4(e): Respondents will not change their estimate of the success of each candidate in future polls based on the sentiment directed at that candidate in the expert chat.

To summarize, we have a hypothesis related to each of the four attributes for which we will examine the effects of social and expert streaming chat: frequency, toxicity, content, and context. Appendix Table A3 displays a summary of these and supplemental hypotheses.

Research design

We conducted a digital field experiment focused on the September 2019 U.S. Democratic Primary debates. Political debates are an exemplar of the kind of political "media event" for which streaming chat is relevant. Primary presidential debates are crucial vehicles for disseminating information. Studies have found primary debates can influence outcomes, such as perceptions of candidate electability (Yawn et al., 1998), affect toward candidates, and issue salience (Best & Hubbard, 1999). Moreover, research on second-screening during political debates has shown the potential for the second-screen to alter the way people process debate information (Bramlett et al., 2018; Jennings et al., 2017), increase engagement with the debate and knowledge acquisition from the debate (Bramlett et al., 2018; Chadwick et al., 2017; Jennings et al., 2020a, 2017; Vaccari et al., 2015) at least for some viewers (McGregor & Mourão, 2017), but also distract and systematically alter the effect of debates on candidate perceptions relative to debate-watching without a second screen (Camaj & Northup, 2019).

The 12 September 2019 Democratic primary debate hosted by ABC News and Univision was the third debate of the 2020 presidential election season, but the first time all the top-polling Democratic candidates would be on one stage. An ABC News press release noted the debate drew 2.9 million unique visitors and 11 million video views (ABC News Live and VOD) across ABC News platforms and partners. Given the amount of interest in the third primary debate and the high level of live-streaming the event attracted, the presidential debate featured in this study serves as an appropriate context for exploring the effects of streaming chat.

We used Amazon MTurk to recruit participants for a two-wave survey. During Wave 1, a Human Intelligence Task (HIT) was made available to all MTurk workers residing in the United States with an approval rating of 95% or higher. Wave 1 included 2,352 respondents. After completing a consent form, respondents who

specified that they were likely to watch the debate, had a Facebook account, and could watch the debate on a computer were deemed eligible and invited to participate in a second survey after the debate.[6]

A total of 1,315 respondents were qualified to participate in the second survey, and 1,095 of these respondents indicated interest. Our sample includes only people who are interested and likely to watch political debates and able to watch online, which means our treatment effects are estimated on a sample that reflects a population that might plausibly be "treated" in this manner during similar events in the real world, an advantage for generalizability.

Experimental conditions

Our experiment is an encouragement design with three conditions. The design is modeled after Gross et al. (2019), who investigate the effect of viewing political debates on different broadcast networks. In our study, respondents who indicated that they were interested in participating in the second survey were randomly assigned and asked to watch the debate on one of the following livestream platforms: FiveThirtyEight (Expert condition), ABC News Facebook page (Social condition), or ABC News (Control condition). Notably, even in the Control condition, subjects are still encouraged to watch the debate. The "control" nature of this condition is the absence of streaming commentary on the assigned platform. For screenshots of the three debate platforms, see Appendix Figure A1.

At the end of the debate, a link to the follow-up Wave 2 survey was emailed to each respondent. Respondents who were eligible and returned for Wave 2 were generally balanced across experimental conditions, suggesting little differential attrition: $N = 305$ in the ABC Control condition (84% recontact rate), $N = 298$ in the Expert condition (82% recontact rate), and $N = 305$ (84% recontact rate) in the Social condition.[7] The analysis sample includes the 576 Democratic respondents (including those who lean toward the Democratic Party) who completed Wave 2 and reported watching at least part of the debate. We focus on Democrats, as they comprise the

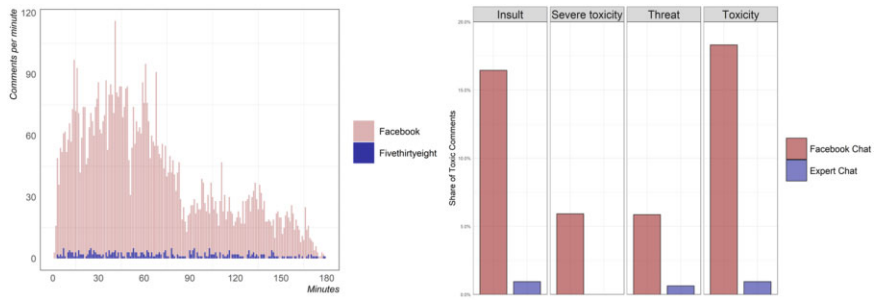

Figure 1 Frequency and toxicity of comments.

substantial majority of the sample, and we anticipated heterogeneous effects by partisanship.[8]

Understanding the chatboxes: Text analysis of the streaming chats

Before presenting our experimental results, we first examine the streaming chats on the ABC News Facebook page and FiveThirtyEight.[9] In the expert chat, there were 300 comments with 2,311 words in total (average 73.8 words per comment). On Facebook, we have a sample of 6,915 comments with 53,092 words (average 7.75 words per comment). Overall, comments in the Social condition are very frequent and short. Users observed 1.68 comments per minute in the expert chat and 38 comments per minute in the social chat. The left plot in Figure 1 presents the evolution of the comments by minute on both platforms. There is a high frequency of comments in the Social condition, especially in the first hour of the debate. In the expert chat, there are comparatively few comments, evenly distributed across the debate.

We now assess the comments in the chat streams, first quantifying the degree of toxicity. We use Google's Perspective application programming interface (API) , a content moderation tool that is an industry standard for automatic detection of toxic content in written comments and which has been used in academic research (Kim et al., 2021; Obadimu et al., 2019; Theocharis et al., 2020). Perspective uses a convolutional neural network model to classify an input text as toxic. The model was built using millions of comments from the internet, using human coders to rate the comments on a scale from "very toxic" to "very healthy."[10] Muddiman et al. (2019) illustrate the limitations of Perspective API (and other pre-trained models) when "they do not have a strong conceptual fit with a concept of interest" (p. 22). Our theoretical interest is in measuring perceived toxicity rather than other related concepts. Therefore, the Perspective API, which was trained on public comment sections, is a useful tool for our case.

Figure 1 (right) presents the proportion for the toxicity scores in four dimensions: toxicity, severe toxicity, threat, and insult. We calculate the proportion of comments in each category that surpass the score of 0.5 in that dimension.[11] Insults and toxic comments appear often in the Facebook chat. In contrast, levels of toxicity in all four dimensions are undetectable in the expert chat, while more than 15% of the comments on Facebook are considered toxic. (See Appendix Figure A2 for the full distribution of toxicity scores by condition.)

Our theory also expects priming effects from the content of the streaming comments. To detect prominent primes in the comments, the research team hand-coded 6,500 comments in the Social condition.[12] First, the research team read all the comments on the Facebook streaming chat with the purpose of classifying when a comment was directed at one of the candidates in the debate. We considered a comment as being directed at the candidates when the text explicitly mentioned the candidate or referred to a policy position or characteristic of the candidate. Three

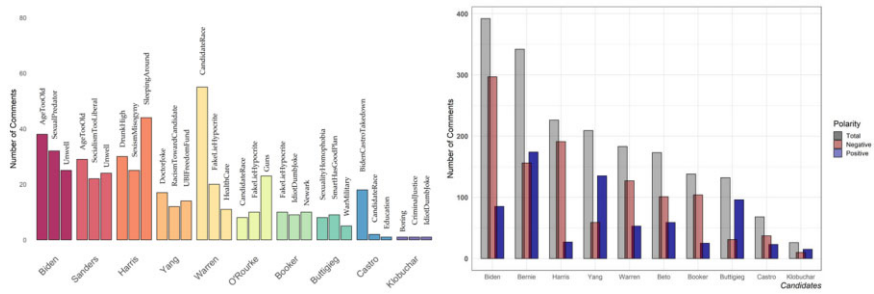

Figure 2 Top three topics and polarity for each candidate.

coders hand-coded all comments, and we used the modal code. Overall, 83% of comments coded by one coder as pertaining to a given candidate saw unanimous agreement.[13] We identified 1,889 candidate-specific comments. We then classified the content of each candidate-specific comment according to topic to understand what was being primed during the debates.[14]

Figure 2 (left) displays the results of our hand-coding of the comment topics. These topics were developed inductively by summarizing common themes in the topics. We did not pre-specify topics we expected to see in the debate comments. Topics include both substantive political issues as well as comments made about candidate personal characteristics. Comments were allowed to be about multiple topics, and if a comment was unclear or its meaning ambiguous, it was left uncoded.[15] Biden was the most frequently mentioned candidate, and the top three topics were all critical: mocking his age (topic labeled "AgeTooOld"); accusing him of being creepy or handsy (topic labeled "SexualPredator"); and suggesting that he was physically unwell. Sanders was also mocked for being old and described as a socialist or too far to the left.

Other notable topics include praise for Yang's UBI proposal and Buttigieg's polished and well-researched plans. There were several mentions of O'Rourke's strong anti-gun statement and Booker's history as mayor of Newark. The comments directed at Harris, however, at times, drifted into territory that was overtly misogynistic and racist, which may be considered outside the bounds of legitimate democratic deliberation. The most common comment accused her of "sleeping her way to the top," in sometimes graphic terms. In addition, the most common single-candidate topic was mockery of Warren by reference to her having claimed Native American heritage. Many of these comments were echoes of Trump's "Pocahontas" moniker.

To conclude, we summarize overall comment "polarity" (whether a comment was positive, negative, or neutral toward the candidate), which we will use to examine *content* and *context* effects.[16] Figure 2 (right) presents our results. The figure displays the total number of comments about each candidate in the Facebook streaming chat and the number of comments with positive and negative valence. Overall, Biden and Sanders received the most attention followed by Harris and

Yang. For most candidates, the comments are mostly negative, with special attention to Harris and Biden, with 85% and 75% of negative comments out of their total share, respectively. Sanders, Yang, Buttigieg, and Klobuchar are the candidates who received proportionally more positive comments, though Klobuchar received very few comments, overall.

Experimental results

We now turn to the results of our experiment by evaluating treatment effects comparing the Social and Expert conditions (with streaming chat) to the Control condition (without streaming chat). We present results for Democratic respondents and their answers to the Wave 2 outcome questions. In each analysis, we report the average effect of assignment to the Expert condition, relative to the Control condition, and assignment to the Social condition, relative to the Control. These are calculated using linear regressions of each outcome on the treatment assignment indicators. The figures to follow present the effect point estimate, as well as 90% and 95% confidence intervals, with two-sided *p*-values in the text.

Our study has a limited amount of two-sided noncompliance (Gerber & Green, 2012) where some subjects in the Control condition who watched the debate may have been "treated" by viewing social or expert commentary, and some subjects in the treatment conditions who watched the debate may have opted not to view the debate on the assigned platform (going untreated). To account for this, we also report complier average causal effects (CACEs) in the Appendix, using two-stage least squares regressions where treatment assignment is used as an instrument for whether the respondent self-reported actually receiving the treatment (i.e., watching the debate on the assigned platform). We find compliance rates of 75% for the Expert condition and 81% for the Social condition. (See Appendix for details.)

Testing hypotheses related to frequency and toxicity effects

Our first set of hypotheses discussed how participants would rate their overall debate experience. Respondents assigned to the Social condition had more negative experiences with the debate than respondents assigned to the Control condition (Figure 3). As expected, respondents who were assigned to watch the debate on the Facebook platform found the debate less informative and enjoyable ($p < .05$) on average, relative to the Control condition without streaming chat.[17] Contrary to our expectations, respondents assigned to the Social condition also found the debate less engaging relative to the Control ($p < .05$).

In addition, contrary to our expectations, assignment to the Expert condition did not have positive effects relative to watching the debate without streaming commentary. Respondents in the Expert condition generally expressed similar, though

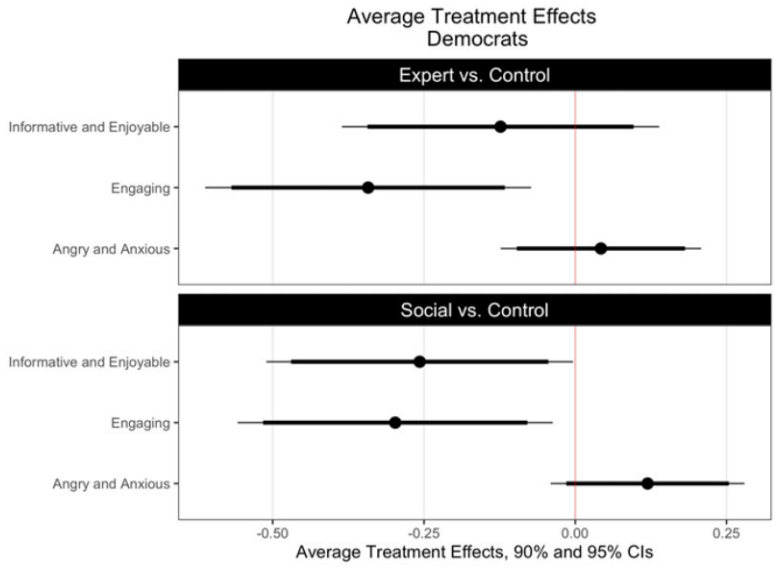

Figure 3 Debate experience.
Notes: Figures display effects comparing average outcomes between the assigned streaming chat condition and the control condition, without streaming chat. Informative, enjoyable, and engaging are 7-point scales, while Angry and Anxious are 4-point scales. See Appendix for question wording.

more muted, reactions. We find no significant differences between the Expert and Control conditions in terms of how informative and enjoyable respondents perceived the debate. Respondents in the Expert condition also found the debate significantly less engaging relative to watching the debate without streaming commentary. Our suspicion is that these Expert results are related to the FiveThirtyEight platform, which provided a very small video window for watching the debate, making it more difficult for participants to engage fully in reading the comments and watching the video (the *broadcast visibility* attribute mentioned above). It is important to remember that our treatment is the entire viewing experience—the platform itself, along with streaming chat.

The text analysis revealed that the comments in the Facebook chat included a great deal of toxicity, aligned with our expectation that watching the debate on Facebook would lead to negative emotional reactions. We find that respondents in the Social condition also reported slightly, but not significantly, higher rates of feeling angry and anxious from watching the debate relative to the Control. Thus, at minimum, the results show that the encouragement to watch the debate on Facebook did not lead debate watchers to have a more satisfying experience and may have led to a less engaging, less enjoyable and informative, and overall negative experience. In Appendix Figures A8–A9, we replicate these results in analyses that

account for compliance. In each case, the results are similar if not stronger for compliers—those who specifically report watching the debate on Facebook or FiveThirtyEight when assigned to do so.[18]

Testing hypotheses related to content effects

We also hypothesized that candidate-specific favorability may be influenced by the presence of consistent negative primes. To assess this, we examine feeling thermometer ratings toward each of the candidates. As shown in Figure 4, respondents in the Social condition came away with more negative perceptions of debate participants, on average, relative to the Control condition. Looking at specific candidates, respondents in the Social condition came away with more negative feelings toward Biden, Harris, and Booker, in particular, as well as O'Rourke and Castro. Notably, these candidates, in particular Biden, Harris, Booker, and O'Rourke, are those who received proportionally more negative comments in our text analysis of the social chat. Results for the Expert condition are slightly more muted, except in the case of evaluations of Castro, which are significantly more negative in the Expert condition.

To provide an intuitive visualization of the close correspondence between the Social chat and the observed effects, Figure 5 plots the average *difference* in the

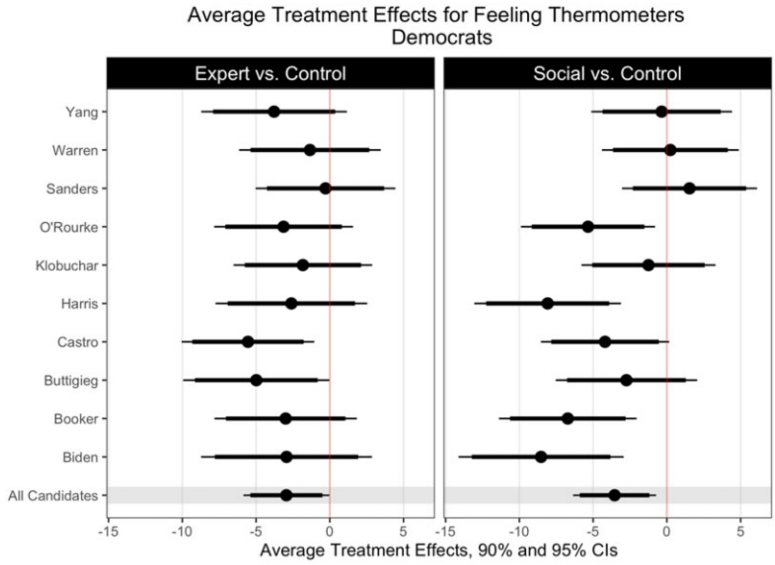

Figure 4 Candidate feeling thermometers.
Notes: Figures display effects comparing average outcomes between the assigned streaming chat condition and the control condition, without streaming chat. Feeling thermometers are on 0- to 100-point scales.

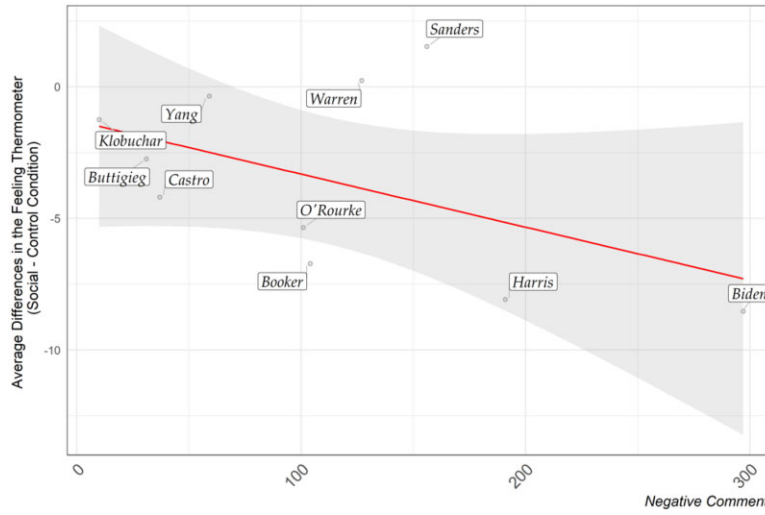

Figure 5 Negative comments decrease candidate feeling thermometers.

feeling thermometer between the Social and the Control condition for each candidate on the sum of comments we labeled as negative on the Facebook streaming chat.[19] The close association is suggestive of the negative content of comments being an important mechanism of the treatment.

Testing hypotheses related to context effects

A final set of hypotheses assessed the extent to which respondents would infer from the commentary how the general public views the candidates. To test this, we asked whether respondents think that the candidate will do better in the polls as a result of the debate. Figure 6 displays the treatment effects of the Social and Expert conditions relative to the Control. While most effects are null, Yang and Sanders were predicted to do significantly better in the polls from those in the Social condition relative to the Control condition. On the other hand, O'Rourke was predicted to do worse in the polls relative to the Control condition.

Similar to the analysis of negative primes, Figure 7 displays the correlation between the number of positive comments a candidate received in the social chat and the difference in projected poll performance between the Social and Control conditions. Notably, Yang and Sanders were candidates who received relatively more positive comments on the ABC News Facebook stream compared to other candidates, and correspondingly, respondents in the Social condition were significantly more likely to predict they would do better in the polls.

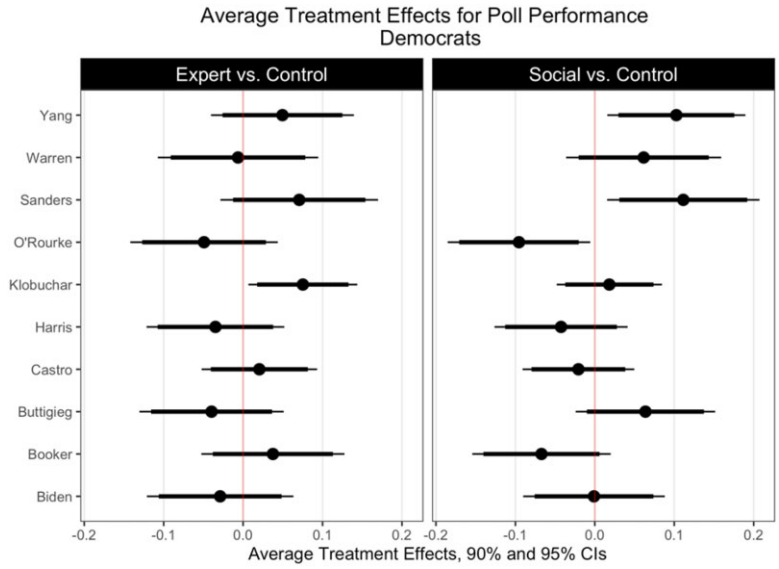

Figure 6 Expected poll performance.
Notes: Figures display effects comparing average outcomes between the assigned streaming chat condition and the control condition, without streaming chat. Poll performance is a 0 or 1 variable, where 1 indicates the respondent believed the candidate would do better in the polls after the debate, 0 otherwise.

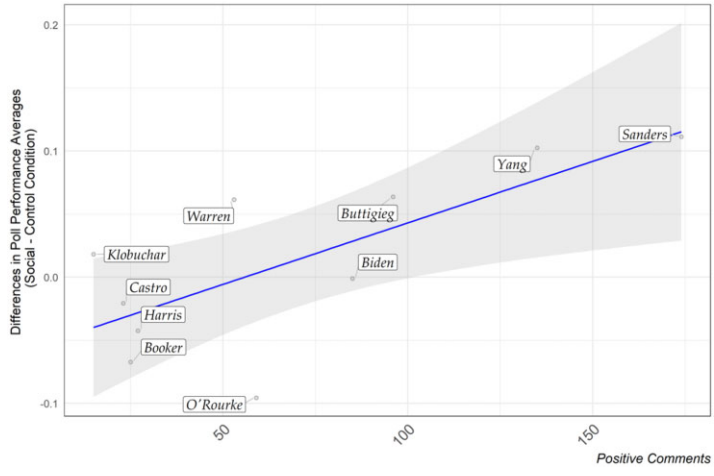

Figure 7 Positive comments increase expected poll performance.

Conclusion

Does streaming chat alter viewer perceptions? We first demonstrate that real-time social commentary can be, in practice, a highly toxic, low-quality, overwhelming, and negative experience that differs greatly from watching an event without comments or with more in-depth and slow-paced expert commentary during a 2019 U.S. Democratic primary debate. We identified four theoretical pathways and hypotheses for which streaming chat may influence the viewing experience: frequency, toxicity, content, and context effects.

First, we anticipated that the introduction of a social chat stream could be distracting for participants due to the quantity and fast-paced nature of the comments. Our text analysis results validated our expectation that the social chat had very frequent comments. Like research showing exposure to a second screen during media consumption can be distracting (Gottfried et al., 2017; Van Cauwenberge et al., 2014), our survey results find that respondents assigned to the social streaming chat came away with less enjoyable, less informative experiences. Thus, similar to media multitasking (Wang & Tchernev, 2012), we find the potential for streaming chat to be less informative, but we do not find that this is offset by respondents feeling the experience was more enjoyable.[20] In addition, while we anticipated the stimulation of the chat could lead participants to feel engaged, respondents in both Social and Expert conditions found the debate significantly less engaging. While we found high toxicity in the Social condition, the Facebook chat made people slightly but not significantly more anxious and angry. Overall, even though streaming chat is growing in usage for major live events, we find no evidence that exposure to streaming chat improved experiences for consumers. If anything, the dynamic but potentially overwhelming and toxic experience of streaming chat led to worse viewing experiences.

Research in the broadcast era identified priming as a key role in the influence of media (Iyengar & Kinder, 1987), including commentary from analysts after political debates (Fridkin et al., 2008) as well as persuasion from the second screen for those who seek it out (Barnidge et al., 2017). We hypothesized that the unique integrated nature of streaming chat would make the content of the chats salient during debates, and influence candidate evaluations above and beyond exposure to just the debate itself in the Control condition. We find support for this hypothesis. Candidates who were subject to high frequencies of negative comments in the social chat, such as Biden, Booker, and Harris, were rated significantly lower in feeling thermometer ratings by respondents in the Social condition relative to the ABC News condition. Lastly, building on Lerman et al.'s (2016) research on the "majority illusion" in social networks, we find evidence that people infer—incorrectly—that the views expressed on the streaming social chat reflect the sentiment of the public. Candidates who received more positive comments (e.g., Sanders and Yang) were predicted to do better in the polls in the Social condition.

While this research design achieves high ecological validity as a digital field experiment involving a major media event, it is not without limitations. One limitation

is the ability to assess compliance. Research has shown that survey respondents report watching presidential debates more often than what administrative data would suggest (Prior, 2012). Though we attempted to minimize incentives to provide misleading compliance information, overstating compliance could understate the effects of streaming commentary on our outcomes. Second, subjects recruited from MTurk are known to be unrepresentative of U.S. adults. That said, Coppock (2019) presents compelling evidence that a wide variety of findings generated from nationally representative samples can be replicated using MTurk. The subjects in our study self-reported already planning to watch the debate. This is not a common attitude, but we argue that it makes them a superior subject pool than a nationally representative sample might.

In addition, while we are confident that streaming chat will continue to be used in future broadcasts, particularly within the political arena (all 2020 U.S. presidential general election debates streamed live on multiple Facebook channels), and in other forms of sports and entertainment, our theoretical framework takes as a premise that the attributes we tested here will vary empirically, according to the nature of the focal broadcast, the composition of the commenters, and the affordances of the platform hosting the streaming chat. Echoing Eveland's (2003) recommendation, we encourage future research in this vein to condition their hypotheses about the "media effects" of a given streaming chat experience on the empirical levels of frequency, toxicity, content, and context. Future designs can assess alternative combinations of these attribute levels, across communication settings, to continue to tease apart their independent effects on viewing experiences. There is also work to be done investigating the attributes (textuality, broadcast visibility, interactivity) that we could not test here.

Outside of streaming chat, the attributes we identify are present to greater or lesser extent in other media and communication experiences. First, "Content is King," as the saying goes. While we certainly agree that the literal *content* will always be an important attribute, Eveland's (2003) concluding admonishment is as applicable today as it was twenty years ago: "Too much research in the current media effects paradigm does not actually address the effects of media. Instead, it focuses on the effects of content that are simply carried by various media" (p. 408).

The dual-screening and media-multitasking literatures that motivated our investigation of *frequency* are themselves responding to the increasing density of stimuli available in the contemporary media environment, a trend we expect to continue—although not necessarily in the form presented by streaming chat. TikTok, for example, accomplishes a novel level of information density by packing many modalities (video, text, music) into a short video.

Toxicity is a widely discussed element of a variety of media experiences, and online discussion platforms—from Reddit to newspaper website comment sections to Twitter conversations—all contend with design choices that can affect the relative degree of toxicity that consumers encounter, such as the degree of anonymity of creators. In the long run, we expect that people with different preferences for toxicity

will self-segregate based on prevailing norms of toxicity; if, however, norms or platform affordances change suddenly, the unexpected presence or absence of toxicity will change the emotional experience for consumers.

We are particularly interested in the evolving role of the *context* attribute, which we think the present experiment is well-scoped to inform. The primary motivation for using social media, Settle (2018) argues, is to learn *social* information: to find out what other people think about other people and events. As social media moves beyond the Facebook model of relying on users' pre-existing social networks (TikTok, YouTube, and Reddit use the algorithmic or social recommendation as their backbone, Auxier & Anderson, 2021), the composition of the public responsible for creating social media content becomes much more difficult to discern. Savvy users may opt into private channels where they can be more confident of the context in which content is created, but the consumers of broader, more anonymous media may make unfounded assumptions about the representativeness of social information to which they are exposed.

Overall, our study points to several consequences of streaming chat under the theoretical conditions outlined. In the context of political debates, streaming chat can shape a more negative and less enjoyable experience that has the potential to spill over into how viewers evaluate candidates and their viability in elections. The study of media effects should account for the pathways by which novel technologies influence the media consumption experience.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Please note: Oxford University Press is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Funding

This research was supported by the Russell Sage Foundation in conjunction with the Summer Institutes in Computational Social Science at Princeton in 2019.

Conflict of interest

The authors affirm that they have no conflicts of interest related to the content of this article.

Notes

1. Facebook reports viewership trends for media https://www.facebook.com/formedia/blog/trends-facebook-live-and-news-publishers.
2. Pre-registration is a practice of registering the research design and hypotheses publicly before starting the process of data collection and analysis. The pre-registration is available through the Center for Open Science (https://osf.io) and is summarized in the Online Appendix.
3. The 20 participants in the FiveThirtyEight chat included a mix of analysts, editors, and writers for the website, several political scientists, and one "Poll Bot." Text of the conversation is available: https://fivethirtyeight.com/live-blog/abc-news-democratic-debate/. The tone of the discourse was conversational and often statistical in nature: referring to polls and embedding graphs and charts related to the topics in the debate.
4. Notably, integrated chat also increases the range of people exposed to streaming chat, as Gil de Zúñiga et al. (2021) demonstrate using WhatsApp.
5. For example, in workplace situations, employees look to low-ranking colleagues for evidence on descriptive norms, not higher-ranking authorities (Dannals et al., 2020).
6. Eligible respondents were offered a bonus payment of $1.50 and an entry into a $100 raffle for their participation in the second survey.
7. Not surprisingly, respondents interested in Wave 2 report more interest in watching the news and are more likely to be Democratic (see Appendix Tables A1 and A2 for sample information).
8. Supplemental results for Republicans are in Appendix (Figures A18–A20).
9. To perform the analysis, we scraped the comments from each of the streaming chats. We were able to extract all the "expert" comments. For Facebook, we were able to extract a large sample of the comments which Facebook makes available through manual search of the chat, by extracting the HTML files of the comments from the public-facing streaming video and chat. We did not collect replies to the comments and reactions from the users.
10. Previous studies comparing manually coded dictionaries and supervised machine learning algorithms to the Perspective API show that the latter exhibited satisfactory accuracy on a variety of tasks (Georgakopoulos et al., 2018; Rajadesingan et al., 2020).
11. The Perspective API is trained as a classification algorithm. Therefore, the scores indicate the probability of a comment being classified as toxic. For this reason, we use the benchmark 0.5 probability cutoff in our analysis, reflecting the threshold above which a person would be more likely than not to perceive a comment as toxic.
12. We hand-coded comments during approximately the first 2.5 hours of the debate. We stopped after a moment when protesters disrupted the debate, and comments described the debate as over.
13. On average, our three coders achieved 84% of agreement on this classification. Our Fleiss' Kappa measure for multiple coders is 0.71, indicating substantial agreement between coders.
14. We provide the same results for the hand-coded comments on the expert chat in Appendix. Supplemental analyses using automated dictionary-based methods also confirm the more negative nature of Facebook comments (Appendix Figure A6).
15. Appendix Table A4 and Figure A3 present a more detailed quantitative analysis of the most frequent words in these hand-labeled topics to provide face validity for the analysis.
16. In the Appendix, we provide a description of our coding rules for the polarity scheme. Our two coders achieved 90% agreement, and a Cohen's Kappa score of 0.81. We use polarity instead of toxicity to investigate content and context effects, recognizing that a

comment does not have to be toxic (rude or disrespectful) in word choice to influence perceptions of candidates. Other forms of negative or positive evaluations may also inform candidate perceptions.

17. We hypothesized the treatments to have similar effects on both how informative and enjoyable respondents perceived the debate. These measures are correlated at 0.7. Given the similar expectation for both measures and this correlation, to reduce measurement error and simplify the presentation, we combined the measures into a single score by taking the average of respondents' answers. We similarly combine anxious and angry into a single measure, following Kim and Kim (2019). Appendix Figure A8 displays results keeping each measure separate.
18. For the Social versus Control comparison, Appendix Figure A9 accounts for an additional form of compliance—whether respondents assigned to the Facebook condition report following along with the Facebook comments. About 70% of respondents report following along in this manner.
19. In Appendix Figure A14, we show that no similar correlation appears in the analysis of the Expert condition. Negative comments are much rarer in this condition.
20. The findings by Wang and Tchernev (2012) might have pointed to a slightly modified version of our H1: both models agree that streaming chat subjects should find the debate less informative, but the media multitasking framework predicts that they should find it more, rather than less, enjoyable, given that multitasking can provide unexpected emotional gratifications. Wang et al. (2012) also demonstrate that the cognitive penalties from media multitasking are reduced if media comes through different modalities: text and audiovisual, in the current case. While our research does not support these hypotheses, they stand as considerations for future research.

Acknowledgments

Author order is alphabetical. We thank the editor, anonymous reviewers, Christopher Bail, Matthew Salganik, and participants at the 2019 Summer Institute in Computational Social Science at Princeton University for their feedback.

References

Anspach, N. M., & Carlson, T. N. (2020). What to believe? Social media commentary and belief in misinformation. *Political Behavior*, *42*(3), 697–718. https://doi.org/10.1007/s11109-018-9515-z.

Arceneaux, K., & Johnson, M. (2013). *Changing minds or changing channels? Partisan news in an age of choice*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226047447.001.00001

Auxier, B., & Anderson, M. (2021). Social media use in 2021. *Pew Research Center*, *1*, 1–4.

Barnidge, M., Gil de Zúñiga, H., & Diehl, T. (2017). Second screening and political persuasion on social media. *Journal of Broadcasting & Electronic Media*, *61*(2), 309–331. https://doi.org/10.1080/08838151.2017.1309416.

Benoit, W. L., Hansen, G. J., & Verser, R. M. (2003). A meta-analysis of the effects of viewing U.S. presidential debates. *Communication Monographs*, *70*(4), 335–350. https://doi.org/10.1080/0363775032000179133.

Benoit, W. L., Pier, P. M., Brazeal, L. M., McHale, J. P., Klyukovski, A., & Airne, D. (2002). The primary decision: A functional analysis of debates in presidential primaries Westport, CT: Praeger. https://books.google.com/books?hl=en&lr=hyFZt3YUqxQC.

Best, S. J., & Hubbard, C. (1999). Maximizing minimal effects: the impact of early primary season debates on voter preferences. *American Politics Quarterly*, *27*(4), 450–467. https://doi.org/10.1177/1532673X99027004004.

Bramlett, J. C., McKinney, M. S., & Warner, B. R. (2018). Processing the political: Presidential primary debate "live-tweeting" as information processing. *An Unprecedented Election: Media, Communication, and the Electorate in the 2016 Campaign*, *169-188*. https://books.google.com/books?id=CtVJDwAAQBAJ.

Camaj, L., & Northup, T. (2019). Dual-screening the candidate image during presidential debates: The moderating role of Twitter and need to evaluate for the effects on candidate perceptions. *Journal of Broadcasting & Electronic Media*, *63*(1), 20–38. https://doi.org/10.1080/08838151.2019.1574117.

Chadwick, A., O'Loughlin, B., & Vaccari, C. (2017). Why people dual screen political debates and why it matters for democratic engagement. *Journal of Broadcasting & Electronic Media*, *61*(2), 220–239. https://doi.org/10.1080/08838151.2017.1309415.

Chen, G. M. (2017). Online Incivility and Public Debate: Nasty Talk. New York: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-56273-5

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. *CSCW: Proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work*, pp. 1217–1230. Portland, Oregon, USA. https://doi.org/10.1145/2998181.2998213

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*, 658–679. https://doi.org/10.1111/jcom.12104.

Coppock, A. (2019). Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods*, *7*(3), 613–628. https://doi.org/10.1017/psrm.2018.10

Dannals, J. E., Reit, E. S., & Miller, D. T. (2020). From whom do we learn group norms? Low-ranking group members are perceived as the best sources. *Organizational Behavior and Human Decision Processes*, *161*, 213–227. https://doi.org/10.1016/j.obhdp.2020.08.002

Eveland Jr, W. P. (2003). A "mix of attributes" approach to the study of media effects and new communication technologies. *Journal of Communication*, *53*(3), 395–410. https://doi.org/10.1111/j.1460-2466.2003.tb02598.x

Fridkin, K. L., Kenney, P. J., Gershon, S. A., & 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. https://doi.org/10.1177/1940161207312677.

Georgakopoulos, S. V., Tasoulis, S. K., Vrahatis, A. G., & Plagianakos, V. P. (2018). Convolutional neural networks for toxic comment classification. *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, pp. 1–6. Patras, Greece. https://doi.org/10.1145/3200947.3208069.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. New York: WW Norton. https://wwnorton.com/books/9780393979954.

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, *12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416.

Gil de Zúñiga, H., Ardèvol-Abreu, A., & Casero-Ripollés, A. (2021). WhatsApp political discussion, conventional participation and activism: Exploring direct, indirect and generational effects. *Information, Communication & Society*, *24*(2), 201–218. https://doi.org/10.1080/1369118X.2019.1642933.

Gil de Zúñiga, H., Garcia-Perdomo, V., & 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. https://doi.org/10.1111/jcom.12174.

Gottfried, J. A., Hardy, B. W., Holbert, R. L., Winneg, K. M., & Jamieson, K. H. (2017). The changing nature of political debate consumption: Social media, multitasking, and knowledge acquisition. *Political Communication*, *34*(2), 172–199. https://doi.org/10.1080/10584609.2016.1154120

Greenberg, J. (2016). Zuckerberg really wants you to stream live video on Facebook. *Wired*. Retrieved from https://www.wired.com/2016/04/facebook-really-really-wants-broadcast-watch-live-video/.

Gross, K., Porter, E., & Wood, T. J. (2019). Identifying media effects through low cost, multiwave field experiments. *Political Communication*, *36*(2), 272–287. https://doi.org/10.1080/10584609.2018.1514447.

Holbrook, T. M. (1999). Political learning from presidential debates. *Political Behavior*, *21*, 67–89. https://doi.org/10.1023/A:1023348513570.

Houston, J. B., McKinney, M. S., Hawthorne, J., & Spialek, M. L. (2013). Frequency of tweeting during presidential debates: Effect on debate attitudes and knowledge. *Communication Studies*, *64*(5), 548–560. https://doi.org/10.1080/10510974.2013.832693.

Iyengar, S. (1987). Television news and citizens' explanations of national affairs. *American Political Science Review*, *81*(3), 815–831. https://doi.org/10.2307/1962678.

Iyengar, S., & Kinder, D. R. (1987). *News that matters: Agenda-setting and priming in a television age*. Chicago: University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/N/bo10579884.html.

Jennings, F. J., Coker, C. R., McKinney, M. S., & Warner, B. R. (2017). Tweeting presidential primary debates: Debate processing through motivated Twitter instruction. *American Behavioral Scientist*, *61*(4), 455–474. https://doi.org/10.1177/0002764217704867.

Jennings, F. J., Bramlett, J. C., McKinney, M. S., & Hardy, M. M. (2020a). Tweeting along partisan lines: Identity-motivated elaboration and presidential debates. *Social Media + Society*, *6*(4), 1-12. https://doi.org/10.1177/2056305120965518.

Jennings, F. J., Warner, B. R., McKinney, M. S., Kearney, C. C., Funk, M. E., & Bramlett, J. C. (2020b). Learning from presidential debates: Who learns the most and why? *Communication Studies*, *71*(5), 896–910. https://doi.org/10.1080/10510974.2020.1807377.

Jungherr, A. (2014). The logic of political coverage on twitter: Temporal dynamics and content. *Journal of Communication*, *64*(2), 239–259. https://doi.org/10.1111/jcom.12087.

Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, *71*(6), 922–946. https://doi.org/ 10.1093/joc/jqab034.

Kim, Y., & Kim, Y. (2019). Incivility on Facebook and political polarization: The mediating role of seeking further comments and negative emotion. *Computers in Human Behavior*, *99*, 219–227. https://doi.org/10.1016/j.chb.2019.05.022.

Larsson, A. O., & Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*, *14*(5), 729–747. https://doi.org/10.1177/1461444811422894

Lerman, K., Yan, X., & Wu, X.-Z. (2016). The "majority illusion" in social networks. *PLoS One*, *11*(2), Article e0147617. https://doi.org/10.1371/journal.pone.0147617.

Masullo Chen, G., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We should not get rid of incivility online. *Social Media+ Society*, *5*(3), Article 2056305119862641. https://doi.org/10.1177/2056305119862641.

McGregor, S. C., & Mourão, R. R. (2017). Second screening Donald trump: Conditional indirect effects on political participation. *Journal of Broadcasting & Electronic Media*, *61*(2), 264–290. https://doi.org/10.1080/08838151.2017.1309418.

McKinney, M. S., & Rill, L. A. (2009). Not your parents' presidential debates: Examining the effects of the CNN/YouTube debates on young citizens' civic engagement. *Communication Studies*, *60*(4), 392–406. https://doi.org/10.1080/10510970903110001.

Muddiman, A., McGregor, S. C., & Stroud, N. J. (2019). (Re) claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, *36*(2), 214–226. https://doi.org/10.1080/10584609.2018.1517843.

Mungeam, F. (2011). *Commenting on the news: How the degree of anonymity affects flaming online*. Gonzaga University, Spokane, WA, USA.

Munger, K. (2019). The limited value of non-replicable field experiments in contexts with low temporal validity. *Social Media+ Society*, *5*(3), 1-4. https://doi.org/10.1177/2056305119859294

Obadimu, A., Mead, E., Hussain, M. N., & 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*, pp. 214–223. Washington, D.C., USA.

Ortoleva, P., & Snowberg, E. (2015). Overconfidence in political behavior. *American Economic Review*, *105*(2), 504–535. https://doi.org/10.1257/aer.20130921.

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283. https://doi.org/10.1177/1461444804041444.

Prior, M. (2012). Who watches presidential debates? Measurement problems in campaign effects research. *Public Opinion Quarterly*, *76*(2), 350–363. https://doi.org/10.1093/poq/nfs019.

Rajadesingan, A., Resnick, P., & Budak, C. (2020). Quick, community-specific learning: How distinctive toxicity norms are maintained in political subreddits. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 557–568. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/7323.

Ran, W., & Yamamoto, M. (2019). Media multitasking, second screening, and political knowledge: task-relevant and task-irrelevant second screening during election news consumption. *Journal of Broadcasting & Electronic Media*, *63*(1), 1–19. https://doi.org/10.1080/08838151.2019.1565659.

Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, *8*(1), 18–33. https://doi.org/10.1080/17512786.2013.813194.

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. New York: Cambridge University Press. https://doi.org/10.1017/9781108560573.

Shah, D. V., Hanna, A., Bucy, E. P., Lassen, D. S., Van Thomme, J., Bialik, K., Pevehouse, J. C. (2016). Dual screening during presidential debates: Political nonverbals and the volume and valence of online expression. *American Behavioral Scientist*, *60*(14), 1816–1843. https://doi.org/10.1177/0002764216676245.

Shin, M., Linke, A., & Kemps, E. (2020). Moderate amounts of media multitasking are associated with optimal task performance and minimal mind wandering. *Computers in Human Behavior*, *111*, 106422. https://doi.org/10.1016/j.chb.2020.106422.

Shockley-Zalabak, P. S., Morreale, S. P., & Stavrositu, C. (2019). Voters' perceptions of trust in 2016 presidential candidates, Clinton and Trump: Exploring the election's outcome. *American Behavioral Scientist*, *63*(7), 856–887. https://doi.org/10.1177/0002764217730855.

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & 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. https://doi.org/10.1111/jcom.12259.

Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The dynamics of political incivility on twitter. *Sage Open*, *10*(2), 1-15. https://doi.org/10.1177/2158244020919447.

Vaccari, C., Chadwick, A., & O'Loughlin, B. (2015). Dual screening the political: Media events, social media, and citizen engagement. *Journal of Communication*, *65*(6), 1041–1061. https://doi.org/10.1111/jcom.12187.

Van Cauwenberge, A., Schaap, G., & Van Roy, R. (2014). TV no longer commands our full attention: Effects of second-screen viewing and task relevance on cognitive load and learning from news. *Computers in Human Behavior*, *38*, 100–109. https://doi.org/10.1016/j.chb.2014.05.021.

Wagner, M., & Morisi, D. (2019). Anxiety, fear, and political decision making. *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228637.013.915.

Wagner, M. C., & Boczkowski, P. J. (2019). Angry, frustrated, and overwhelmed: The emotional experience of consuming news about president trump. *Journalism*. https://doi.org/10.1177/1464884919878545.

Warner, B. R., Carlin, D. B., Winfrey, K., Schnoebelen, J., & Trosanovski, M. (2011). Will the "real" candidates for president and vice president please stand up? 2008 pre-and post- debate viewer perceptions of candidate image. *American Behavioral Scientist*, *55*(3), 232–252. https://doi.org/10.1177/0002764210392160.

Wang, Z., David, P., Srivastava, J., Powers, S., Brady, C., D'Angelo, J., & Moreland, J. (2012). Behavioral performance and visual attention in communication multitasking: A comparison between instant messaging and online voice chat. *Computers in Human Behavior*, *28*(3), 968–975. https://doi.org/10.1016/j.chb.2011.12.018.

Wang, Z., & Tchernev, J. M. (2012). The "myth" of media multitasking: Reciprocal dynamics of media multitasking, personal needs, and gratifications. *Journal of Communication*, *62*(3), 493–513. https://doi.org/10.1111/j.1460-2466.2012.01641.x.

Weaver, D., & Drew, D. (1995). Voter learning in the 1992 presidential election: Did the "nontraditional" media and debates matter? *Journalism & Mass Communication Quarterly*, *72*(1), 7–17. https://doi.org/10.1177/107769909507200102.

Yawn, M., Ellsworth, K., Beatty, B., & Kahn, K. F. (1998). How a presidential primary debate changed attitudes of audience members. *Political Behavior*, *20*(2), 155–181. https://doi.org/10.1023/A:1024832830083.

Zhu, J. H., Milavsky, J. R., & Biswas, R. (1994). Do televised debates affect image perception more than issue knowledge? A study of the first 1992 presidential debate. *Human Communication Research*, *20*(3), 302–333. https://doi.org/10.1111/j.1468-2958.1994.tb00325.x.