Disclosure statement

5 March 2025

I, Lehan Zhang, hereby declare that I have no relevant material or financial interests that relate to research described in the article entitled “Adding Fuel to the (Gun)Fire: How Politicians Polarize the Public Debate.”

Sincerely,



Lehan Zhang
PhD Candidate,
ETH Zurich

Disclosure Statement

Author: Gabriele Gratton
Submission Title: Adding Fuel to the (Gun)Fire: How Politicians Polarize the Public Debate

6 March 2025

I, Gabriele Gratton, hereby declare that I have no relevant material or financial interests that relate to the research described in the article "Adding Fuel to the (Gun)Fire: How Politicians Polarize the Public Debate"



Gabriele Gratton
Professor of Politics and Economics
UNSW Business School

Disclosure statement

5 March 2025

I, Pauline Grosjean, hereby declare that I have no relevant material or financial interests that relate to research described in the article entitled “Adding Fuel to the (Gun)Fire: How Politicians Polarize the Public Debate.”

Sincerely,

Pauline Grosjean

Pauline Grosjean
Professor of Economics,
UNSW Sydney

Disclosure statement

6 March 2025

I, Hasin Yousaf, hereby declare that I have no relevant material or financial interests that relate to research described in the article entitled “Adding Fuel to the (Gun)Fire: How Politicians Polarize the Public Debate.”

Sincerely,



Hasin Yousaf
Senior Lecturer, School of Economics,
UNSW Sydney

Adding Fuel to the (Gun)Fire: How Politicians Polarize the Public Debate*

Lehan Zhang† Gabriele Gratton‡ Pauline Grosjean§
Hasin Yousaf¶

March 8, 2025

Abstract

We show how politicians polarize policy-relevant public debates. Analyzing 4.75 million tweets about 57 mass shootings events, two-way fixed effects and event studies results show that the partisan and tribal content of tweets systematically increases after a politician first tweets about an event. An analysis of the timing of interventions suggests that politicians do not intervene in response to specific characteristics of the event, nor to immediate changes in the debate. Interventions by other (not politicians) focal influencers do not change the debate. We show how the rhetorical supply of politicians explains the unique polarizing effect of political interventions.

Keywords Twitter, Partisanship, Polarization, Tribalism

JEL codes D72, L82, P00, Z13

*We are grateful to Ruben Durante, Sophie Hatte, Federica Izzo, Federico Masera, Adam Zelizer, Katia Zhuravskaya and audiences at the Econometric Society Australasian Meetings, ENS Lyon, the Australian Political Economy Workshop, and ETH Zurich for insightful comments and discussions. We thank Henrik Osterberg and Yang Wu for excellent research assistance. Gratton and Grosjean are recipients of Australian Research Council Future Fellowships (FT210100176 and FT190100298). Gabriele Gratton's research was supported under the Australian Research Council's Discovery Projects funding scheme (Project DP240103257). This project received IRB approval from UNSW (approval number HC220333).

†ETH Zurich. Email: lehan.zhang@gess.ethz.ch.
‡UNSW. Email: g.gratton@unsw.edu.au.
§UNSW and CEPR. Email: p.grosjean@unsw.edu.au.
¶UNSW. Email: h.yousaf@unsw.edu.au.

1 Introduction

American contemporary public debate is characterized by increasing divisiveness, partisanship, and tribalism. Online as well as offline, citizens increasingly describe the other side of the argument as the enemy—a threat to their own identity (e.g., Abramowitz and Webster, 2016; Finkel et al., 2020; Iyengar et al., 2019). Scholars and pundits alike have raised concerns that such polarization between partisan and tribal groups may pose a threat to the functioning and sustainability of democracy (Finkel et al., 2020; Levitsky and Ziblatt, 2019; Sunstein, 2009).

A growing literature has pointed to the catalyzing effect of social media in amplifying political divisions and undermining social cohesion (Enikolopov et al., 2024; Guriev et al., 2023; Manacorda et al., 2022, 2023; see Zhuravskaya et al., 2020 and Campante et al., 2023 for reviews). Consequences can be tragic when online hate fuels physical violence and hate crimes (Bursztyn et al., 2019; Müller and Schwarz, 2021).[1] However, there is still substantial debate about the drivers of such online polarization. The literature so far has stressed how social media enable demand-side factors that explain partisanship and tribalism: contact with users who share beliefs and world views reinforces extreme views (Allcott and Gentzkow, 2017; Bakshy et al., 2015; Levy and Razin, 2019; Levy, 2021; Nyhan et al., 2023; Sunstein, 2009); anonymity disinhibits social users (Ederer et al., 2023); and users engage more with more controversial and hateful speech (Beknazar-Yuzbashev et al., 2022; Giavazzi et al., 2024; Guess et al., 2023).[2] According to these explanations, the role played by political elites is at most marginal, with the possible exception of particularly famous and disruptive personalities, such as Donald Trump (Bursztyn et al., 2020; Grosjean et al., 2023; Müller and Schwarz, 2023).

Several features of the political and media environments suggest that the literature may have overlooked the role of supply-side factors in spurring political and social divisions. Political elites are highly polarized[3] "influencers"—many with more than one million Twitter followers[4]—able to amplify news, generate new narratives, and sway voters

[1] See also Adena et al. (2015) for related evidence on the effect of more traditional media (radio) on hate crimes.

[2] Recent experimental studies on Facebook and Instagram document substantial levels of partisan segregation online (González-Bailón et al., 2023) but found that algorithms and consumption of like-minded news sources do not affect polarization in the short run (Nyhan et al., 2023). However, the null result on polarization has come under recent scrutiny, when it was revealed that Facebook may have manipulated the default algorithm during the experiment (Thorp and Vinson, 2024).

[3] Barber and McCarty (2015), Fowler et al. (2022), and Moskowitz et al. (2024), among others, point out that political elites are more polarized along party lines than the general population of voters.

[4] As we document, many Congress members and state governors have more than one million Twitter followers; ex-presidents Barack Obama and Donald Trump are in the top 10 worldwide rank of accounts by number of followers.

(Boffa et al., 2024). It is then natural to wonder what role politicians may play in fueling online polarization.

In this paper, we document how political communication on social media steers online public debates towards more partisan and tribal expressions. In particular, we show that policy relevant public debates, such as those surrounding mass shooting events, become more partisan and tribal after politicians intervene in the conversation. We further establish the unique effect of interventions by politicians, as opposed to interventions by other public figures with similar or even larger online reach and exposure, such as singers, actors, sports stars, or tycoons. The specific effect of political communication is partly due to politicians *supplying* partisan rhetoric, a pattern of communication that we do not observe either for other elites or for traditional news media organizations.

To show this, we collect data on the universe of 4.75 million tweets related to the 57 mass shooting events covered on the first page of the *New York Times* between March 2016 (when Twitter rolled out its algorithmic timeline) and October 2022 (when Elon Musk acquired the company) and we identify changes in the language of tweets around the time when politicians intervene in these online debates. We focus on mass shootings for several reasons. First, mass shootings are particularly salient news events, which polarize lawmakers and voters not only on gun and crime control policies (Yousaf, 2021), but also on unrelated issues, including social and environmental policies (Barilari, 2024).[5] Second, the nature of mass shootings, with a well-identified start time, makes it possible to pin down exactly when the debate begins online and the first related tweet by a politician. Third, their unanticipated nature guarantees that politicians or political institutions had no influence on the onset of the debate.[6]

For each mass shooting event, we collect all related tweets, constructing a longitudinal "public debate" dataset, timed from the start of the event to seven days after the first related tweet.[7] Within each event's public debate, we identify a *political intervention* as a tweet originating from the account of a U.S. politician with more than one million Twitter followers (70 prominent politicians).[8] We also track interventions by other influencers who are not politicians, identified from a list of top 100 Twitter accounts by number of

[5]This partisan polarization systematically results in legislative gridlock (McCarty et al., 2016), possibly fueling a spiral towards greater and more toxic polarization (Jacob et al., 2024).

[6]In contrast, major natural disasters such as hurricanes have no clear start time and the onset of the debate surrounding them is directly affected by political and bureaucratic institutions monitoring potentially dangerous tropical cyclones.

[7]As discussed in Section 2, the volume of tweets spikes in the first hour and a half after the start of an event and remains very stable after three days, so that our choice of stopping the data collection seven days after the event is innocuous.

[8]We show that our results do not change if we focus on the much larger sets of 117 U.S. politicians with more than 500,000 followers or 202 U.S. politicians with more than 200,000 followers.

followers.[9]

For every tweet in our dataset, we use dictionary-based methods to measure the tweet's *partisanship* and *tribalism*. We measure a tweet's partisanship by whether the tweet contains words in the dictionary of partisanship derived by Gentzkow et al. (2019) from Congressional speech. We measure tribalism by whether the tweet contains words that express loyalty and betrayal, as defined by the Moral Foundations Dictionary (Enke, 2020; Graham et al., 2009).[10] We then adopt event-study and two-way fixed effects methodologies at the tweet-event level, controlling for event-specific and time-since-onset of the debate fixed effects to measure if, and how, a political intervention changes expressions of partisanship and tribalism in the public debate—our measure of online polarization.

Our first contribution is to describe when and how politicians intervene in public debates. We uncover substantial variation in the identity and partisanship of the politicians who intervene in the public debate. In total, 67 (out of 70) different politicians intervene in 52 of the 57 public debates, with 29 distinct politicians intervening first in a debate.

There is also substantial variation in the timing of political interventions, both with respect to the timeline and characteristics of the events themselves and to the timing of interventions by other politicians, news media organizations, and other influencers. Political interventions do *not* arrive systematically immediately after the start or the end (i.e., incapacitation of the shooter) of an event, nor do they immediately follow other political interventions. Traditional news media organizations tweet about all events and do so very quickly after the start of the event. Political interventions occur, in general, well after interventions by news media organizations, and without any systematic pattern. Politicians also do not systematically follow other influencers, who tend to intervene at a later stage of the public debate compared to politicians. We find that both the timing and specific wording of political interventions, in terms of partisanship or tribalism, are also uncorrelated with events' characteristics, such as location, victim count, or race of the shooter. Last, with respect to debate characteristics themselves, we do not observe that politicians systematically time their intervention with the peak of the debate in terms of volume of total related tweets, or in terms of the popularity (followership, likes) of users or tweets. In fact, political interventions tend to arrive well after the debate has peaked.

Event study analyses further show that the timing of the first political intervention is also unrelated with immediate *changes* in both the partisan and tribal content of the debate as well as in the volume of tweets. Such a lack of systematic patterns in the data

[9] We exclude associations and organizations, such as sports clubs and news outlets, as well as U.S. and non-U.S. politicians. This leaves us with 62 individual accounts.

[10] See `https://moralfoundations.org` (retrieved December 9, 2023).

suggests that politicians do not (or are unable to) systematically time their interventions in response to changes in the partisan or tribal content of the public debate.

Our second contribution is to document whether and how politicians polarize the public debate. We show that mass shooting events become topics of heated partisan and tribal discussions *after* politicians intervene in the public debate. Our two-way fixed effects and event study analyses compare the share of partisan and tribal content in a debate after the first political intervention with the share of partisan and tribal content in the debate before the intervention and in other debates in the same time interval since their respective onsets. Event-specific fixed effects account for any potential heterogeneity across events, which may be systematically correlated with the polarizing nature of the event. Because the nature of the debate may also naturally change over time, as emotions get heated up or die off, in ways that could systematically co-vary with political interventions, we control for a set of fixed effects for highly granular time intervals since the onset of the debate,[11,12] as well as for a time trend since the onset of the debate, which—in robustness checks—we allow to vary across events.

Our descriptive analysis of political interventions, showing an absence of systematic patterns in the timing of political interventions, as well as our event studies results, showing the absence of pre-trends in partisanship, tribalism, and volume of tweets prior to the first intervention, suggest that politicians do not strategically time their interventions, lending credence to a causal interpretation of our results. Our results show that political interventions trigger an immediate and persistent increase in partisanship in online public debates. The magnitude of the effect is substantial. In the first hour after the first political intervention, the probability that a given tweet contains partisan language is 43% greater than in the hour before the intervention. By two hours after the intervention, it is 88% greater. The effect on tribalism is slower to materialize and amounts to a 24% increase relative to the pre-intervention mean in the 60 to 120 minutes after the intervention. Intensive margin estimates show that subsequent political interventions further polarize the debate, but do so at a decreasing rate.

By contrast, interventions by other influencers who are not politicians do not polarize the public debate. To understand the specific effect of political interventions, we first document that politicians substantially differ in their rhetoric from other social media users. In particular, *no* intervention by a non-political influencer or news media organization

[11]Our baseline specification accounts for 30 minutes fixed effects, and we show that our results are robust to shorter or longer intervals.

[12]Our setting is not suitable to the use of calendar time fixed effects, since events occur at different dates with little overlap in the 7-days time frame after an event. Calendar time fixed effects (hourly, weekly, monthly) would be redundant with our 30 minutes time interval fixed effects.

contains partisan language, while 10 out of 52 (first) political interventions do. Politicians are also much more likely to use tribal language compared with news media organizations (44% vs. 5%). Other elites sometimes use tribal rhetoric, but only when (and after) politicians do so.

We then show that these differences in the rhetorical supply are crucial to understand the polarizing effect of political interventions. Political interventions that are partisan are associated with a substantially larger increase in the polarization of the debate. Overall, this evidence suggests that political interventions polarize the public debate in part because they inject polarizing rhetoric in the debate. We also show that our results are primarily driven by the effect of political interventions on the content of tweets posted by newcomers to the debate, who only tweet about the event after the first political intervention, and are stronger and accelerated when events take place outside regular working hours and on weekends. Nevertheless, we show that political interventions do not attract more users to the debate nor change the type of users—in particular, they do not attract more politicized users. These results are consistent with the view that political interventions that inject partisan rhetoric in the public debate place mass-shooting events under a polarized lens that shapes the language of users who join the debate after the intervention takes place.

This paper contributes to the literature on media and political polarization. A large literature has established how traditional media affect political participation (Angelucci et al., 2024; Lenz and Lawson, 2011; Wang, 2023) and political preferences (e.g., Couttenier et al., 2024; Wang, 2021b; Zhuravskaya et al., 2020). Access to slanted news and exposure to focal cable news and radio figures can deeply transform political preferences and significantly sway election results (Adena et al., 2015; Amarasinghe and Raschky, 2022; Ash et al., 2024; DellaVigna and Kaplan, 2007; Martin and Yurukoglu, 2017; Wang, 2021a), even to the point of fostering ethnic hatred (DellaVigna et al., 2014) and pushing people to perpetrate genocide (Yanagizawa-Drott, 2014). Several studies have focused on specific aspects of the online media environment to better understand the additional or differential effects of social media over traditional media. In particular, the literature has highlighted how homophily in online networks may give rise to segregated news environments—so called "echo chambers"—that may reinforce partisanship (within chambers) and amplify division and polarization (across chambers) (Gillani et al., 2018; Levy and Razin, 2019). However, recent experimental studies suggest that the polarizing effects of echo chambers may have been overestimated (Guess et al., 2023; Nyhan et al., 2023). Moreover, while segregation in online media environments is indeed higher than for traditional media and exacerbated by platforms' algorithms (González-Bailón et al., 2023; Guess et al.,

2023), it is lower than in face-to-face interactions (Gentzkow and Shapiro, 2011). Other studies have focused on how online social media affect the wider media landscape and the provision of information (Cagé et al., 2022; Hatte et al., 2021), with a particular focus on the veracity and reliability of information (Allcott and Gentzkow, 2017; Guess et al., 2020). A widespread concern is that social media facilitate the spread of rumors and misinformation, which might distort users' political preferences and sway their votes towards extreme policy platforms. This concern has spurred major policy debate about the implementation and enforcement of fact-checking and censorship (Guriev et al., 2023; Henry et al., 2022; Jackson et al., 2022; Mattozzi et al., 2023). However, recent evidence suggests that users actually consume relatively little political or elections news on social media (Allcott and Gentzkow, 2017) and that a majority are able to discern false from true information (Angelucci and Prat, 2024). In this paper, we shift away from the debate on misinformation on political or election-related news. In contrast, our focus is on public debates over major policy-relevant news events that do not directly originate in any politician's decision. More generally, we shift away from the debate on how the environment generated by social media may amplify users' behavioral biases and focus instead on how the platform aspect of social media, where focal influencers can directly communicate with a large number of users, may affect expressions of partisanship and tribalism and thereby accentuate political and social divisions.[13]

The remainder of the paper is as follows. Section 2 describes our data. In Section 3, we document patterns and trends (or the absence thereof) in the timing of political interventions. In Section 4, we motivate and describe our empirical strategy. In Section 5, we present and discuss our main results. In Section 6, we investigate the mechanisms through which political interventions polarize public debates and provide evidence on the role of the supply of partisan rhetoric in shaping the public debate. We conclude in Section 7.

2 Data

We now discuss the construction of our database of online public debates and our outcome measures of partisanship and tribalism. Appendix C includes more details.

[13] Acemoglu et al. (2010) present a model of misinformation spread where "forceful" agents have an asymmetric effect on other users' beliefs. Our focus differs from theirs in that we do not study beliefs and misinformation but rather expressions of political and social divisions.

2.1 Mass shooting events

Mass shooting events have been shown to affect polarization among both voters and lawmakers (Barilari, 2024; Yousaf, 2021). We aim to uncover whether political communication on social media contributes to this polarization. From a measurement point of view, mass shooting events offer several advantages. They provide a homogeneous set of events that: (i) generate a high volume of tweets, so that we can measure variation in language; (ii) have a well-defined and unanticipated starting time, so that we can identify the first tweet (by any user) and the first tweet by a politician related to the event; and (iii) are not directly determined (in their timing, location, and specific characteristics) by the actions or decisions of politicians. We consequently focus on the 57 mass shooting events that appear on the front page of *The New York Times*[14] between March 1, 2016 and October 27, 2022.[15] For each event, we retrieve: (i) the *start time*—the start of the shooting episode; and (ii) the *end time*—the time at which the shooter is either arrested or dead (on average, 116 minutes after the start time). We also retain several characteristics of the event location characteristics (a school in 21% of the events; a business in 19%), the race of the shooter (white in 51% of the events), and the number of victims (on average, 7.7).

2.2 Tweets

For each event, we search for all related tweets following a methodology similar to Demszky et al. (2019). For each event, we define two sets of keywords, one related to the circumstances of the event and one related to its geographic location. A tweet is related to that event if it contains at least one word from each set of keywords.[16]

For each tweet related to an event, we retrieve: the exact time of posting, the text of the tweet, the tweet ID, and information about the posting account, including: username, user handle, profile description, and user ID. Figure 1 shows an example of the raw data retrieved from a tweet and posting account's profile.

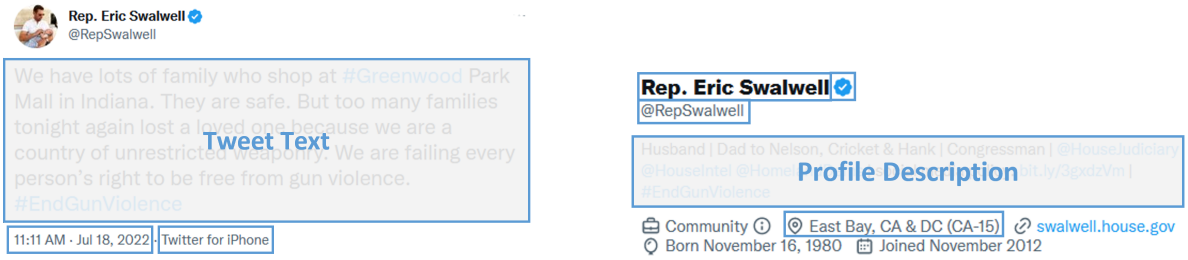
2.3 Public debates

For each event, we identify the first related tweet and all related tweets posted within the following seven days. We refer to this sequence of tweets as the event's *public debate* and

[14] We extract this information from *The New York Times* archive `https://archive.nytimes.com/`

[15] For the time period 2021–2022, we check the overlap between events on the front page of the *The New York Times* and events on the front page of other popular publications (*The Washington Post*, *The Los Angeles Times*, *The Wall Street Journal*, and *USA Today*).

[16] Appendix C contains more details on the selection process including the list of keywords.



(a) Data retrieved from tweet: exact time of posting, text of the tweet, tweet ID

(b) Data retrieved from posting account: username, handle, and profile description

Figure 1: Example of data retrieved for each tweet

we define the *onset* of the public debate as the time of the first tweet. We exclude retweets and non-English language tweets. We retain a total of 4.75 million tweets ($89{,}442$ tweets per debate, on average).

Figure A.1 shows the evolution of (the average) public debate over time in the first 24 hours (Panel (a)) and over seven days (Panel (b)). The Figure plots the distribution over time of all tweets in a public debate, aggregated in 10 minutes windows. Less than 0.7% of the public debate occurs in the first 30 minutes. The volume of tweets then rapidly picks up and peaks, on average, around 80 minutes after the onset of the debate. It then decreases steadily from its peak until 15 hours after its onset. The volume of the debate keeps decreasing and stabilizes at a very small level after two and a half days (at around 0.05% of all tweets in the debate per 10 minutes window). Overall, 60% of the public debate occurs in the first 24 hours and 87% in the first 72 hours. Given the low volume of tweets after two and a half days, our choice of stopping tweet extraction after seven days is likely inconsequential.

2.4 Measures of polarization of the public debate

We measure the online polarization of a public debate with the amount of partisan and tribal language used in its tweets. We measure partisanship and tribalism with dictionary-based methods. We treat the text of each tweet i as a "bag of words" or "bag of phrases," $\mathcal{B}_i$, where each word or phrase[17] in the tweet is an element of the set $\mathcal{B}_i$. For each outcome y (partisanship or tribalism), its dictionary is a set $\mathcal{D}_y$ of words or phrases. We measure tweet i's outcome y_i as equal to 100 if $\mathcal{B}_i \cap \mathcal{D}_y \neq \emptyset$ and 0 otherwise. Dictionary-based methods are standard in the social science literature (Gentzkow et al., 2019) and provide several advantages for our analysis. They are transparent, provide tweet-level estimates

[17] As we detail below, some of our dictionaries are made of single words. Others are made of longer phrases, containing 2 to 5 words.

that are stable to variations in the underlying samples and corpora, and have a straightforward interpretation as the probability that any given tweet contains words from a given dictionary.

Partisanship. We measure partisanship by whether the tweet contains phrases from a dictionary of 78 partisan phrases identified by Gentzkow et al. (2019) (the complete list is in Appendix C.0.1). This list includes the top 10 two-word most Republican and top 10 two-word most Democratic phrases per each Congressional Session from 2005 through 2016. It is compiled with methods aimed at selecting phrases that most enable "an observer to infer a congressperson's party from a single utterance" (Gentzkow et al., 2019).

Tribalism. We measure tribalism by whether the tweet contains words from the dictionary of 52 words expressing loyalty and betrayal, as defined by the Moral Foundations Dictionary created by Graham et al. (2009) and used by, e.g. Enke (2020) (the complete list is in Appendix C.0.2). Expressions of loyalty and betrayal capture "people's emphasis on being loyal to the in-group" (e.g., family or political party,) (Enke, 2020) and are "related to our long history as tribal creatures".[18] They are not specific to any political party and not predictive of partisan identity (Koleva et al., 2012). Expressions of loyalty and betrayal have been linked to the contemporary American "culture wars" (Enke, 2020; Graham et al., 2009; Haidt and Graham, 2007; Koleva et al., 2012).

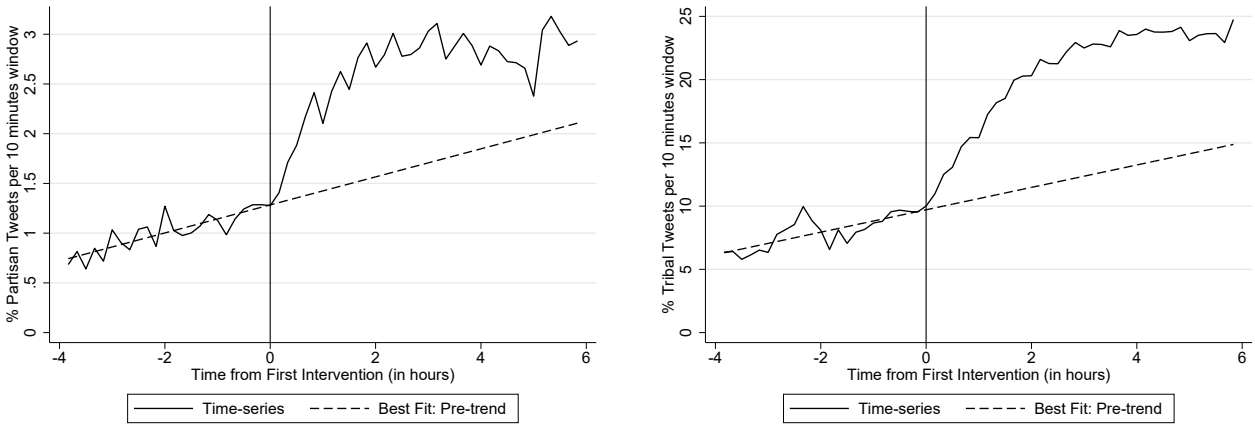
2.5 Interventions

Political interventions. To identify when politicians *intervene* in a public debate, we compile a list of all U.S. politicians with more than one million Twitter followers as of July 11, 2022.[19] These accounts are selected from a list of all past and present U.S. presidents and members of Congress, presidential candidates, state governors, and mayors of the top 50 cities by population, between 2016 and 2022. This search returns 70 Twitter accounts (listed in Appendix C, Table C.3).

For each public debate, we identify the first and subsequent tweets originating from the accounts of these politicians. We call these tweets the *political interventions* in the public debate. 52 out of 57 public debates feature a political intervention. The 52 first interventions originate from 29 different accounts. In 73% of cases, the first political intervention is by a Democrat. The politicians who most frequently intervene first in a debate

[18] See `https://moralfoundations.org` (retrieved December 9, 2023).

[19] We consider the robustness of our results to using alternative cutoffs of 500,000 or 200,000 followers.



(a) Partisanship (b) Tribalism

Figure 2: First political intervention, partisanship and tribalism: raw time series

These figures plot the percentage of tweets containing partisan (panel (a)) or tribal content (panel (b)) averaged over 10 minutes intervals, for events that receive a political intervention. The vertical line indicates the first political intervention in the debate.

are Rep. (D-CA) Ted W. Lieu (6 first interventions), Rep. (D-CA) Nancy Pelosi (4), and Rep. (D-CA) Eric Swalwell (4).

Most debates have interventions from multiple politicians. The median number of political interventions is 7, and the mean is 10.[20] The politicians who intervene most overall are President Donald Trump (42 total interventions), the New York City Mayor[21] (41), and Sen. (D-NY) Chuck Schumer (39).

Figure 2 illustrates the core message of the paper: the partisan and tribal content of the public debate disproportionately increases after the first political intervention, relative to a simple time trend fitted to the pre-intervention period. A natural concern, however, is that the timing of the political intervention may systematically coincide with specific junctures in the event itself, or be strategically timed with changes in the debate. We discuss in greater detail the patterns and timing of political interventions in Section 3 and implement two-way fixed effects and event-study methodologies (decribed in Section 4) in Sections 5 and 6.

Interventions by other influencers. We construct a comparable list of prominent influencers who are not politicians. Many thousands of athletes, artists, and business celebrities have more than 1 million followers, and some of them are very active in public de-

[20] On average, a debate has 7.6 (median: 5) interventions by Democratic politicians and 2.5 (median: 1) by Republicans.

[21] This account appeared under the name Bill de Blasio (D) until December 31, 2021 and Eric Adams (D) from January 1, 2022

bates. We retrieve a list of the top 100 accounts by followers (all have in excess of 25 million followers) and eliminate the accounts of politicians (including non-U.S. and non-active politicians) and organizations, leaving us with 61 individuals (listed in Appendix C.0.3 Table C.4).[22]

3 How politicians intervene in the public debate

In this section, we describe the patterns of political interventions in the public debate. First, we uncover substantial variation in the timing of the first political intervention. Second, we note the lack of systematic patterns in the timing of political interventions with respect to one another or with respect to interventions by other focal users or traditional news media organizations. Third, we show that the timing of political interventions is not systematically associated with the characteristics of the event itself, nor with characteristics of the public debate, in terms of partisanship, tribalism, or volume of tweets. In total, these observations motivate our empirical strategy (described in Section 4) to identify the effect of political interventions on the public debate.

3.1 Timing of interventions

First political intervention. We have already discussed the extent of variation in the identity of the first politician who intervenes in a debate, with 29 distinct politicians being first to intervene in a debate. Our data also reveal great variation in the timing of political interventions. The median time of the first political intervention (the vertical solid line in Figure A.1) is 135 minutes after the onset of the debate; the mean (vertical dashed line in Figure A.1) is 226 minutes. Both are well after the onset and the peak of the debate (around 80 minutes after the onset; see Section 2).

Appendix Figure A.2a reveals great variability in the timing of the first political intervention with respect to key event and debate timelines. The earliest (first) intervention occurs 42 minutes after the start of the event (30 minutes after the onset of the debate); the latest as much as 1431 minutes (almost 24 hours) after the start of the event. The 25th percentile of the distribution of the timing of the first intervention is at 106 minutes after the start of the event and the 75th percentile is more than 4.4 hours after. In only a handful of events does the first intervention occur before the end of the mass shooting

[22]The final list includes a number of celebrities that are influential only among U.S. minorities, as well as all the most influential celebrities from a variety of backgrounds, such as Elon Musk, Justin Bieber, Rihanna, Taylor Swift, Ellen De Genres, or Bill Gates.

event, with the median intervention occurring 114 minutes minutes later. The figure also shows that *all* political interventions occur after the first intervention by a news media organizations[23] (90 minutes later, at the median), but politicians first intervene in the debate well before other influencers: there is a 235 minutes median delay between the first tweet by a politician and the first tweet by another influencer, with a very wide dispersion (interquartile range of 572 minutes).

Subsequent political interventions. Most events in which a politician intervenes receive multiple political interventions (10 on average). Four debates receive a single political intervention, and 42% have 4 or fewer political interventions (see Figure A.3 in Appendix). Only 18% of the events have 20 or more political interventions, with the maximum (89) reached for the Orlando nightclub shooting in 2016. The multiplicity of political interventions raises the question of whether politicians strategically react to one another.

If politicians strategically monitored and responded to one other, we would expect subsequent interventions to immediately follow the first one, and to display much reduced variability in their timing compared with the first intervention. Statistics displayed in Figure A.2b for the first 10 political interventions show that this is not the case. The interquartile range of the second, third, ..., and up to sixth intervention is comparable to that of the first one, and is even larger for subsequent interventions. There are also substantial delays between interventions, with the median second intervention occurring 205 minutes after the onset of the debate (i.e., as much as 70 minutes after the median first intervention). While the evidence so far suggests little or no strategic timing of political interventions with respect to one another or to key events in the timeline of the event, we now turn to more formal tests of these relationships.

3.2 Event and debate characteristics and political interventions

Table 1 shows how the timing of the first political intervention correlates with a broad set of characteristics of: (i) the intervening politician, including: gender, party affiliation, follower count, and the partisan and tribal content of their tweet (Panel A); (ii) the event itself, including: race of the shooter, location of the shooting, and number of casualties (Panel B); and (iii) the public debate, including: volume of tweets, likes, retweets, and in-

[23]We use an approach similar to the approach used to select politicians' and other elites' accounts to identify 76 accounts of the most influential English-language news organizations covering U.S. events on Twitter. Appendix C.0.3 Table C.5 contains the full list of media organizations accounts. This list includes, e.g., @cnnbrk, @cnn, @nytimes, @FoxNews, and @Reuters.

dicators of follower count of the twitter users engaged in the debate (Panel C). Columns 1 and 2 display the mean and standard deviation of the timing of the first political intervention for different values of each characteristic. Out of the 18 characteristics of the politician, event, and debate, only one—the party affiliation of the politician—is associated with a statistically significant difference in the timing of the intervention (as shown in Column 3), which is less than one would expect at random. In particular, and most important for our analysis in the following sections, the timing of the intervention is unrelated to the partisan or tribal content of the intervention tweet.

Table B.1 in Appendix conducts a similar analysis to explore whether the *number* of political interventions depends on the politician, event, and debate characteristics. The number of interventions covaries significantly with only four out of 18 characteristics. There are substantially more interventions in the events with higher victim count, when the shooter is not Black, and when the first politician to intervene is a male or Republican. However, the partisan and tribal content of the first political intervention, the online popularity of the debate, and the follower count of the first politician to intervene are uncorrelated with the number of subsequent interventions. In our empirical investigation, we account for event fixed effects, which account for any heterogeneity in event characteristics that could attract more politicians to tweet as well as for any characteristic of the first politician that intervenes and other fixed characteristic of the first political intervention.

Overall, the analysis of the timing of the first and subsequent interventions and the absence of any systematic differences between the timing or number of interventions and debate characteristics suggest that politicians do not strategically time their intervention with observable characteristics of the online debate . In Section 5, we use an event-study methodology (described in Section 4) to further show that political interventions are also unrelated to immediate *changes* in the characteristics of the online debate.

4 Empirical strategy

The discussion and results in Section 3 underpin our strategy. To fix ideas, one way to think of the process generating an intervention by a specific politician is as one that includes deterministic, stochastic, and idiosyncratic components. The start of a public debate triggers a stochastic process that informs politicians of the event. While all politicians eventually become informed, the exact time at which politician P becomes informed has a stochastic component that depends on her individual and event-specific characteristics, but also on unrelated variables such as P's schedule of travels and business on the day

Table 1: Timing of political intervention

Dependent Variable (DV): Intervention Time			
	Mean of DV when:		
Independent Variable (X)	X=0	X=1	Difference
Panel A: Politician characteristics			
Male	198.66	239.49	40.82
	(152.35)	(276.25)	(59.49)
Republican	266.79	104.19	-162.61***
	(264.60)	(74.63)	(47.23)
SUPPLY: Partisan	221.89	244.01	22.13
	(255.97)	(180.97)	(68.19)
SUPPLY: Tribal	251.15	194.61	-56.53
	(301.21)	(135.47)	(62.73)
Followers Count: High	240.91	192.92	-47.98
	(273.72)	(149.95)	(58.94)
Panel B: Event characteristics			
Shooter race: White	222.52	210.50	-12.03
	(204.52)	(269.3)	(67.02)
Shooter race: Black	206.35	249.24	42.88
	(247.19)	(223.76)	(76.81)
Shooter race: Other	221.42	195.81	-25.61
	(254.98)	(190.27)	(69.34)
Shooting location: School	209.55	239.39	29.84
	(172.52)	(381.74)	(111.17)
Shooting location: Business	227.24	166.75	-60.49
	(254.33)	(88.64)	(48.71)
Shooting location: Community	226.68	162.20	-64.48
	(251.72)	(86.89)	(48.55)
Shooting deaths: High	237.93	201.88	-36.05
	(277.39)	(148.77)	(59.12)
Shooting length: Long	187.18	298.46	111.28
	(232.66)	(228.81)	(85.52)
Panel C: Twitter characteristics before intervention			
Tweets Volume: High	223.00	233.89	10.89
	(266.43)	(174.94)	(62.62)
Poster's Followers: High	228.01	223.15	-4.86
	(282.26)	(164.63)	(62.03)
Poster's Followings: High	233.41	213.51	-19.90
	(279.38)	(163.75)	(61.45)
Event's Like Count: High	217.60	245.35	27.75
	(267.64)	(176.92)	(62.6)
Event's Retweet Count: High	228.39	222.82	-5.57
	(285.16)	(165.11)	(62.66)

The unit of observation is an event. Column 1 (respectively, 2) shows the mean and standard deviation of the time (in minutes from the onset of the debate) of the first political intervention when the variable in the corresponding row is equal to 0 (respectively, 1). Column 3 displays the coefficients estimated from separate OLS regressions (with robust standard errors) of the timing of the first political intervention on each row variable.

of the event. From the moment politician P becomes informed, she may, depending on politician-event-specific characteristics, begin the process of formulating her intervention. Once the politician has formulated an intervention, then a stochastic process pins down the exact time when she or her staff are able to enact the intervention (post the tweet). The exact delay between becoming informed and tweeting is therefore also in part stochastic and varies across politicians and events. The stochastic nature of these delays is what generates the quasi-random nature of the timing of political interventions we documented in Section 3 and allows us to study the effect of such interventions on the public debate.

We conduct our analysis at the tweet-event level. We normalize time within each public debate so that the onset of the public debate is time 0. Therefore, each tweet in a debate is posted at a time t from 0 to T (7 days later). Our objective is to estimate whether a political intervention in a public debate changes the probabilities that subsequent tweets in the same debate contain partisan or tribal language. To do so, we begin by estimating the effect of the *first* political intervention.

For each tweet i posted at time $t \in [0, T]$ in event e's debate, we define $D_{i,e,t}^{(u,v)}$ as a dummy variable equal to 1 if and only if $t \in (u, v)$ (i.e., if the tweet is posted between times u and v). Let p_e be the time of the first political intervention in event e's debate.

4.1 Two-way fixed effect specifications

Our baseline two-way fixed effect analysis estimates:

$$y_{i,e,t} = \beta D_{i,e,t}^{(p_e,T)} + \eta_e + \theta_t + \zeta t + \upsilon_{i,e,t} \tag{1}$$

where $y_{i,e,t}$ is, alternatively, whether tweet i contains expressions of partisanship or tribalism, η_e is an event-specific fixed effect, θ_t denotes a set of 30-minutes intervals fixed effects, and ζ captures the influence of time trends, measured in continuous time since the onset of the event. In this specification, the pre-intervention period $(0, p_e)$ is the excluded reference window. We address inference issues with the estimation of two-way fixed effect models (Borusyak et al., 2024; De Chaisemartin and d'Haultfoeuille, 2020; Sun and Abraham, 2021) in Section 5.4.

The nature of the debate, including in terms of partisanship and tribalism, and the characteristics of the political intervention may vary across events, for example depending on victim count, or location. Although, as we discussed, the timing of the intervention does not systematically vary with event characteristics, we still include event-specific fixed effects, η_e. These fixed effects account for any unobserved heterogeneity in event-specific debate or political interventions. The dispersion of tweet content may also vary

greatly across events. For this reason, we cluster standard errors by event.

Another immediate concern is that the nature of the public debate naturally changes over time. These changes could stem both from a change in the nature of the debate (for example, tweets may organically become more or less partisan as users see more arguments brought forward) or from a compositional change among twitter users, as different people join or exit the debate as time goes by. For example, if more partisan users become interested in the event as the debate gains track, the debate may naturally become more partisan over time. We account for these potential changes in the debate using intervals of time-since-the-onset-of-the-debate fixed effects. In (1), θ_t denotes a set of 30-minutes intervals fixed effects from the start of the event. We consider alternative intervals of 15 or 60 minutes in robustness analysis. We also allow a linear time trend to account for continuous changes in the debate over time.

Because the volume of tweets varies over time, the dispersion of the error term may also vary over time, which may affect the precision of our estimates. We therefore cluster standard errors over the time intervals dimension, in addition to the event dimension. These standard errors adjustments correct inference issues related to the serial correlation of the error term $\upsilon_{i,e,t}$ within an event over time, and within intervals of time since debate onset, across events.

In further robustness checks, we account for the fact that users posting immediately after a political intervention may not have had time to read the intervention, or may have formulated their tweet before the intervention and experienced a small delay in posting it, so that the tweet appears in our dataset *after* the intervention but could not have been affected by it. To account for this mis-classification risk, we include a separate dummy $D_{i,e,t}(p_e, p_e + \epsilon)$ for tweets posted immediately after the intervention and show that our result is robust to the inclusion of this dummy for both ϵ equal 1 and 2 minutes.

A potential limitation of specification (1) is that, as noted in Section 2, political interventions arrive at different times between time 0 and T for different events. This implies that the pre-intervention period, $(0, p_e)$, and the post-intervention period, (p_e, T), are heterogeneous across events, so that β averages across different time periods, which may complicate the interpretation of the results.

To harmonize comparison windows, we estimate, for $a, b, c > 0$,

$$y_{i,e,t} = \alpha D_{i,e,t}^{(0,p_e-b)} + \beta_1 D_{i,e,t}^{(p_e,p_e+a)} + \beta_2 D_{i,e,t}^{(p_e+a,p_e+c)} + \beta_3 D_{i,e,t}^{(p_e+c,T)} + \eta_e + \theta_t + \zeta t + \upsilon_{i,e,t} \qquad (2)$$

$D_{i,e,t}^{(0,p_e-b)}$ captures all tweets from the onset of the debate until b minutes *before* the intervention; $D_{i,e,t}^{(p_e,p_e+a)}$ captures all tweets in the a minutes *after* the intervention; $D_{i,e,t}^{(p_e+a,p_e+c)}$

captures all tweets between a and c minutes after the intervention; and $D_{i,e,t}^{(p_e+c,T)}$ captures all tweets posted *later* than c minutes after the intervention. The coefficients of interest, β_1 and β_2, capture the changes in the partisan or tribal content of tweets posted in the a minutes after the intervention (short-term), and between minutes a and c after the intervention (medium-term), compared to the comparison window of length b before the intervention. The coefficient β_3, which captures this change later than c minutes after the intervention until T (long-term) is difficult to interpret because, as time goes on in the debate, further interventions within the debate and other news events may influence the rhetoric of the public debate. Based on the event-studies analysis of the dynamics of the effect, we set a, b and c to 60 minutes.

In the classical potential outcome framework, the assumption required to causally identify the effect of a political intervention is that, in the absence of a political intervention at time p_e, after controlling for the time elapsed since the onset of the debate, as well as both time intervals since the onset of the debate and event fixed effects, the partisan and tribal content of the public debate would remain unchanged (Equation (1)), or at least would remain unchanged between b minutes before the intervention and $a+c$ minutes after (Equation (2)). This assumption would be violated if politicians systematically timed their interventions in response to the changes in public debate we hope to characterize. We address potential violations of the identification assumption in several ways. First, the high-dimensional time fixed effects ensure that we are comparing tweets posted after a political intervention (treated tweets) to tweets posted in a debate about a similar event but in which a politician has not intervened (yet, or at all) within the same (short) interval of time after the onset of the debate. The inclusion of event-specific time trends in our robustness analysis further enables us to account for different linear dynamics in the nature of tweets across debates. Second, we have already discussed how the timing of an intervention is uncorrelated with any of the debate characteristics, including its prominence (volume of tweets, number of likes and retweets) or the profile of users engaged in it. Moreover, we begin our analysis with an event-study analysis that shows that there are no pre-trend differences in the partisan or tribal content of the debate, as well as in the volume of public debates, before a political intervention.

4.2 Event-Study specification

An event study analysis offers several advantages over the two-way fixed effect approach described thus far. First, it allows us to test for pre-trends in outcomes before a political intervention. Second, it enables us to study how political interventions affect the public

debate over time, rather than averaging over the whole window from the time of the intervention to either the whole post-intervention period or a, or $a + c$ minutes later, and to estimate precisely when the effect of a political intervention materializes (thereby also motivating our choice of parameters a and c). To smooth out noise in daily observations, we estimate parameters for 10-minutes bins, starting from 180 minutes before the political intervention and ending 300 minutes after it. Therefore, the event-study specification estimates (where time is expressed in minutes):

$$y_{i,e,t} = \alpha D_{i,e,t}^{(0,p_e-180)} + \sum_{\tau=-18}^{-2} \beta_\tau D_{i,e,t}^{(p_e+10(\tau),p_e+10(\tau+1))} + \sum_{\tau=0}^{29} \beta_\tau D_{i,e,t}^{(p_e+10\tau,p_e+10(\tau+1))} + \beta_T D_{i,e,t}^{(p_e+300,T)} + \eta_e + \theta_t + \zeta t + \upsilon_{i,e,t}. \tag{3}$$

As in (2), we include fixed effects for each 30-minutes interval from the onset of the debate, event fixed effects, and a linear time trend. The omitted time bin includes all tweets posted in the 10 minutes prior to the political intervention (i.e., in the interval $(p_e - 10, p_e)$).[24]

4.3 Marginal effects of political interventions

So far, we focused on estimating the effect of the first political intervention in the public debate. Most public debates receive multiple interventions. Therefore, we estimate the following equation to capture the marginal effect of subsequent political interventions:

$$y_{i,e,t} = \gamma f(\sum_{0}^{t} D_{i,e,t}^{(0,t)}) + \eta_e + \theta_t + \zeta t + \upsilon_{i,e,t}. \tag{4}$$

For each tweet i posted at time t, the treatment of interest is the sum of political interventions that have occurred until t.[25] We characterize f by including higher order polynomials of the number of cumulative interventions and plotting the marginal effect of each intervention.

[24] Borusyak et al. (2024) show that in some settings with staggered treatment where each unit is treated only once, econometric models with unit and time fixed effects are unable to identify a unit-specific linear trend. This happens in fully dynamic settings where all units are eventually treated. Our framework is different because our estimation sample includes events that are never treated, in the sense that no politician posts a tweet related to these events.

[25] For instance, suppose that a public debate receives two political interventions, one 60 minutes into the debate, and the other 90 minutes into the debate. Then, the variable $\sum_0^t D_{i,e,t}^{(0,t)}$ takes value 0 for $t < 60$, 1 between $t = 60$ and $t = 90$, and 2 for $t > 90$.

5 Political interventions polarize public debates

We now show that political interventions increase the partisan and tribal content of the public debate. The first political intervention has the largest effect, but subsequent interventions polarize the debate further, albeit at a decreasing rate. We also discuss the robustness of our results to alternative specifications, as well as additional results.

5.1 Event-study results

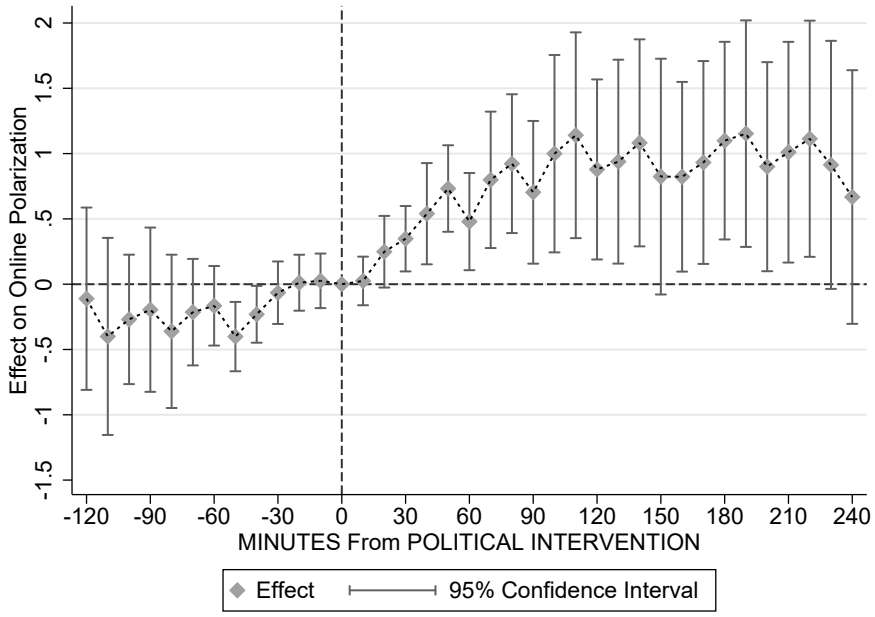
Figure 3 displays the estimates of β_τ, $\tau \in \{-12, \ldots, 24\}$ in (3) for partisan (Panel (a)) and tribal (Panel (b)) language. The results show that the timing of the first political intervention is unrelated to prior trends in the partisan or tribal content of the public debate. Yet, the probability that a tweet in the public debate contains partisan or tribal language significantly increases after the first political intervention.

The effect of the first political intervention on partisanship is immediate. The probability that a tweet contains partisan language is $.25$ percentage points higher in the 10–20 minutes window after the intervention, compared to before. The effect builds up, increasing to $.54$, $.80$, and 1.00 percentage points in the 30–40, 60–70, and 90–100 minutes windows after the intervention. After 90 minutes, the effect remains stable, between $.8$ and 1.1 percentage points, and persists at this level in the following two and a half hours (up to 240 minutes).

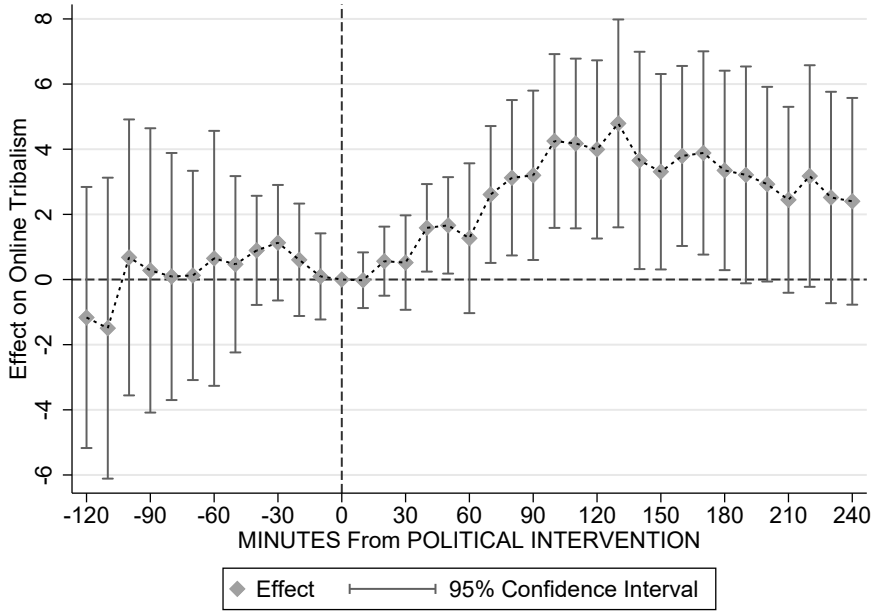
In comparison, the effect of the political intervention on tribalism takes more time to materialize and is less persistent over time. The probability that a tweet contains tribal language is 1.59 percentage points higher in the 30-40 minutes window after the intervention, compared to before, an effect that is statistically significant at the 5% level. The effect increases steadily, to 2.61, 4.25, and 4.79 percentage points in, respectively, the 60-70, 90-100, and 120-130 minutes windows after the intervention. The effect then gradually attenuates and is no longer statistically significant after three hours.

5.2 Two-way fixed effects results

Column 1 in the first panel of Table 2 displays the estimate of β from Equation (1) for partisanship content of a tweet in the 7 days after the first political intervention in the debate. The probability that a tweet contains partisan language increases by $.81$ percentage points (significant at the 1% level), which represents a 75% increase relative to the pre-intervention mean.



(a) Partisanship



(b) Tribalism

Figure 3: Event-Study Results

These figures plot OLS coefficients with 95% confidence intervals. The plotted coefficients are the β_τ coefficients associated with each 10 minute window, described in Equation (3). The outcome variable is Partisanship in panel (a) and Tribalism in panel (b). Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate.

The second panel of Table 2 displays the estimates associated with Equation (2). The probability that a tweet contains partisan language increases by .47 percentage points

Table 2: First political intervention, partisanship and tribalism

VARIABLES	(1) Partisanship	(2) Tribalism
Equation 1		
POST-Intervention (β)	0.810***	2.136
	(0.278)	(1.415)
Observations	4,747,621	4,747,621
R-squared	0.014	0.025
Time Trend	Linear	Linear
Equation 2		
POST-Intervention$< 1h(\beta_1)$	0.465***	0.564
	(0.120)	(0.803)
POST-Intervention$1 - 2h(\beta_2)$	0.958***	3.203***
	(0.303)	(1.151)
POST-Intervention$> 2h(\beta_3)$	0.983***	2.271
	(0.360)	(1.501)
Observations	4,747,621	4,747,621
R-squared	0.014	0.025
Time Trend	Linear	Linear
Mean DV	1.08	8.90

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (column 1), or tribal language (column 2). Panel A shows the OLS estimates of β from Equation (1). Panel B shows the OLS estimates of β_1, β_2, and β_3 from Equation (2) with $a = b = c = 60$. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

(a 43% increase relative to the pre-intervention mean) in the first 60 minutes after the first political intervention, compared to the 60 minutes before the intervention, and by $.96$ percentage points (an 88% increase relative to the pre-intervention mean) in the 60-120 minutes after the intervention. Both coefficients are statistically significant at the 1% level. The effect of a political intervention on partisanship is long-lasting: partisanship is $.98$ percentage points higher more that two hours after the political intervention relative to 60 minutes before.[26]

Column 2 displays the corresponding estimates for tribalism. The probability that a

[26]The coefficients estimated from (1) and (2) are not directly comparable to one another because the excluded comparison windows are different.

Table 3: Number of political interventions, partisanship and tribalism

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism	(5) Partisanship	(6) Tribalism
# Interventions	0.108***	0.451*				
	(0.031)	(0.249)				
# Interventions Squared	-0.004***	-0.011				
	(0.001)	(0.008)				
Log(# Interventions)			0.250*	1.864***	0.053***	0.181**
			(0.142)	(0.420)	(0.017)	(0.073)
Observations	4,747,621	4,747,621	4,748,530	4,748,530	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025	0.014	0.025
Weights	None	None	None	None	Followers	Followers

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (column 1), or tribal language (column 2). The table shows the OLS estimates of β from Equation (4) for different functional forms of γ. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

tweet contains tribal language increases by 2.14 percentage points after the first political intervention, a 24% increase relative to the pre-intervention mean (first panel). Although the coefficient is not statistically significant at the 10% level (p-value: $.13$) over the whole post-intervention window, the second panel of Table 2 shows that this masks heterogeneity over time, with a short-lived, yet statistically significant, 3.2 percentage points (a 36% increase relative to the pre-intervention mean) increase in tribalism in the 60-120 minutes after the intervention.

Overall, these results align with the dynamics uncovered in the event study figure: an immediate, large, and persistent effect of a political intervention on partisanship, and a slower, less persistent—and yet non negligible—effect on tribalism in the medium term.

5.3 Marginal effects of political interventions

Table 3 displays the estimates of γ from Equation (4) for different specifications of f. In Columns 1 and 2, we estimate a quadratic relationship between the number of political interventions and our outcomes of interest. The estimated coefficients show that more political interventions have a larger polarizing effect on the public debate, but at a decreasing rate.

Appendix Figure A.4 plots the marginal effects of an OLS estimation of partisan or tribal content of a tweet posted at time t as a quartic function of the number of political interventions that have occurred until t. The first few interventions have the strongest impact on the probability that the tweet contains partisan (Panel a) or tribal (Panel b)

language. The marginal effect of an additional intervention by a politician becomes statistically indistinguishable from zero after the 7th intervention for partisanship and the 12th intervention for tribalism.

Given the non linear relationship between polarization and the number of interventions, we specify f as a logarithm function for ease of interpretation. Estimates in Columns 3 and 4 of Table 3 suggest that each doubling of the number of interventions results in a $.25$ percentage points increase in the partisan content of a tweet and a 1.86 percentage points increase in the tribal content of a tweet.

To address the possibility that the popularity of intervening politicians on social media may amplify their polarizing effects, which implies that our model, which considers a simple sum of political interventions, may be misspecified, we estimate a weighted version of (4), in which we weigh the number of interventions by the cumulative number of followers of the politicians that have intervened until time t.[27] This additional estimation, in Column 5 for partisanship and Column 6 for tribalism, shows that our results are robust.

5.4 Additional results and robustness

Volume of Tweets. We estimate our event study specification with the volume of tweets in each 10-min window as the dependent variable. The results displayed in Figure 4 show the absence of any pre-trend in the volume of the debate prior to a political intervention, lending further credence to the main identifying assumption that politicians do not systematically time their intervention in response to immediate changes in the debate. The results also shows that the volume of tweets is unaffected by a political intervention, suggesting that political interventions do not radically change the online debate beyond its partisan and tribal content.

Long-term effects. The results are consistent in event-studies that consider a longer time horizon of 15 hours after the intervention. These results are displayed in Appendix Figure A.5, with one-hour bins to ease visualization. Partisanship increases steadily after the first political intervention, stabilizes after 4 hours, and remains significant in the long run. Tribalism only materializes one to two hours after the intervention, increases in the first 4 hours, but declines over time and is no linger significant 7 hours after the intervention.

[27] For instance, suppose there is an event with only two interventions with the first intervention occurring at $t = 60$ and second at $t = 90$. Suppose the politicians who intervenes first has 1.2 million followers, while second politician has 3.4 million followers, then our main independent variable takes value 0 for $t < 60$, 1.2×10^6 for $60 < t < 90$, and 4.6×10^6 for $t > 90$.

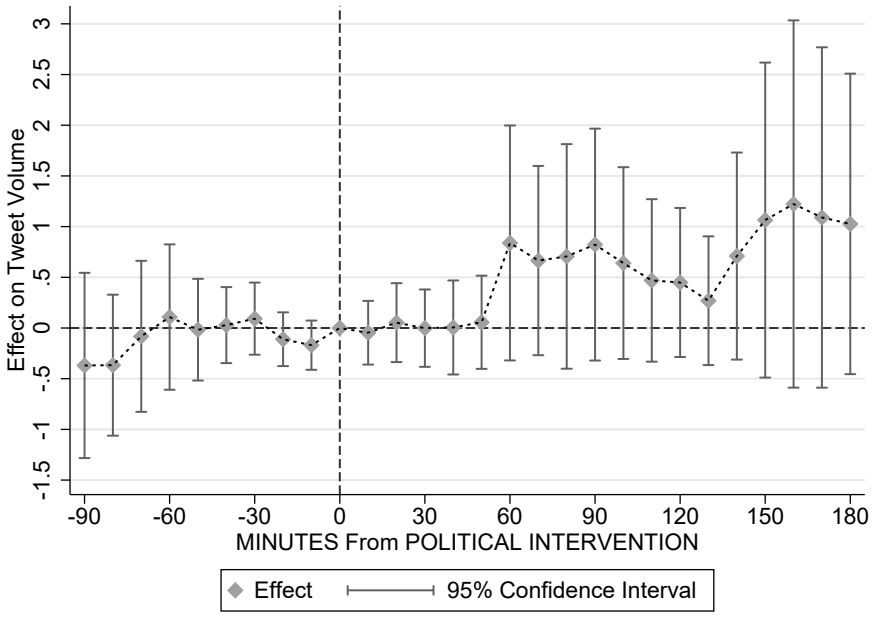


Figure 4: First Political Intervention and Volume of Tweets

The figure plots the OLS coefficients with 95% confidence intervals. The plotted coefficients are the β_τ coefficients described in Equation (3). The dependent variable is the number of tweets posted in each 10 minute window. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate.

Robustness of event-study and two-way fixed effect results. Recent literature on staggered DID highlights potential issues with the two-way fixed effect estimator stemming from the fact that the estimated parameter is a weighted average of each treatment (in our context, each political intervention) where the weights may be negative. We follow the recommended diagnostic by De Chaisemartin and d'Haultfoeuille (2020) and compute the weights associated with each treatment.[28] Figure A.6 shows that there is little variation in the weights and that only a handful (less than 10%) is negative. We also address possible inference issues related to the simultaneous inclusion of time trends and event-fixed effects in the event study (Borusyak et al., 2024). Appendix Figure A.7 shows that the results are unchanged when we exclude the time trend from Equation (3). Finally, we apply the method proposed by Sun and Abraham (2021) to estimate the treatment effect for each event individually. For partisanship, the mean and median treatment effects are 1.00 and 0.72, respectively. For tribalism, the mean and median treatment effects are 2.25 and 3.09. This shows that the median and average of individual treatment effects are very similar to the main estimates obtained using the two-way fixed effect methodology, suggesting little heterogeneity in treatment effects.

We implement sensitivity analyses to potential violations of the parallel trends as-

[28]We collapse our data into 30-minute windows for estimation and concentrate on the short and medium-term effects.

sumption. The canonical concern here is that unit-specific time shocks may affect the outcome even in the absence of the treatment. In our setting, since treatment occurs at different moments in calendar time, unobserved time shocks that may systematically affect treated and untreated events differently (related, e.g., to the news of the day) are not of concern. A concern would arise in our setting if interventions always occurred at the same time (since the onset of the event), which may systematically coincide with key junctures in the debate that may systematically affect outcomes. We have discussed in Section 3 that this is not the case. We still implement the sensitivity analysis suggested by Roth (2022) and Rambachan and Roth (2023). The results, displayed in Appendix Figure A.8, show that the post-treatment violation of parallel trends should be more than 1.5 to 2 times the *maximum* pre-treatment violation in order to explain away our results. Given the granularity and high dimensionality of the data used to estimate the trends, based on tweets in real time, such large deviations seem unlikely.

To ensure that Equations (1) and (2) are capturing the effect of political interventions rather than artifacts generated by the specification, we show the results of permutation inference tests (based on $1{,}000$ replications) in which we randomly assign the timing of political interventions across events. We consider different restrictions in the timing of political interventions, to occur either at any time in the post period (left panel), to mirror our estimate in Equation (1)), or during the time intervals considered in Equation (2) (middle and right panels). The results in Figure A.9 consistently show that our effect sizes are well outside the range of estimated effects from these placebo treatments. Our randomization inference results are not centered around 0 because our sample consists of a higher proportion of post-treatment observations relative to the pre-treatment sample (the post-treatment sample is 93.2% of the overall sample).[29]

We implement a series of additional robustness checks. First, we check that our results are unchanged when we correct for potential missclassification of tweets posted immediately after the first political intervention. To do so, we exclude tweets posted within one or two minutes after the treatment. The results are unchanged (Appendix Table B.2). Second, we allow the time trend since onset of the debate to vary across events by adding an event-specific time trend to our estimation. Results, displayed in Appendix Table B.3, are robust. Third, we show that our results are insensitive to varying the definition of the time dimension of the fixed effects by considering shorter (15 minutes) or longer (60 minutes) time intervals since the onset of the debate (Table B.4). Fourth, we show in Appendix Ta-

[29]This means that we over-sample from the post-treatment sample in randomized inference, thus resulting in our estimates being centered around a small positive number (0.06 for partisanship for estimates of β_1, and 0.10 and 0.05 for partisanship and tribalism estimates of β_2; β is centered around 0).

ble B.5 that our results are robust when we change the set of politicians considered in the analysis. We augment our sample to include all the (first) interventions by any politician with at least $500{,}000$ (Columns 1 and 2) or $200{,}000$ (Columns 3 and 4) followers. Our two-way fixed effects estimates are consistent with our baseline results, suggesting that our results are not due to selective inclusion of specific politicians in our analysis set. Last, we explore the sensitivity of our results to any particular event. Figure A.10 in Appendix displays estimates of β, β_1, and β_2 in a series of estimations of Equation (1) and (2) when we exclude one event at a time. The coefficients are stable and similar to our main estimates across specifications.

6 How political interventions polarize public debates

This section investigates the mechanisms through which political interventions polarize online public debates. We start by establishing that interventions by politicians shape the public debate in systematically different ways when compared to interventions by other public figures. We then delve into the specificity of political interventions, and show how the rhetorical supply by politicians contributes to online polarization. Last, we show that our results are mostly due to newcomers to the debate, rather than a change in the expression of users already engaged in the debate.

6.1 Non-political interventions do *not* polarize public debates

The effects of political interventions we have documented so far may simply be due to the salience, centrality, and popularity of politicians online. Politicians are focus users with large follower counts and broad reach (Alatas et al., 2019; Bakshy et al., 2011), and their followers are more likely to consume and diffuse their messaging (Anger and Kittl, 2011). Given ideological segregation over news consumption on the internet (see, e.g., Guess et al., 2023; Levy and Razin, 2019; Levy, 2021), the effects we observe following a political intervention may not be due to the political intervention itself, but to the news diffusion and amplification that an intervention by a focal user entails. Another possibility is that politicians are divisive by their very nature, so that their interventions naturally divide users across camps.

To investigate these possibilities, we perform the same analysis for similarly salient, central and public figures, but who are not politicians. We select these infuencers through a selection procedure as close as possible to the one used for politicians (see Section 2.5). These public figures intervene in 19% of the debates. They tend to have many more fol-

Table 4: Non-political interventions, partisanship and tribalism

VARIABLES	(1) Partisanship	(2) Tribalism
POST-NON-POLITICAL (β)	-0.301	-0.028
	(0.395)	(1.104)
Observations	4,748,501	4,748,501
R-squared	0.014	0.025
Time Trend	Linear	Linear
Equation	1	1

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (Columns 1 and 3), or tribal language (Columns 2 and 4). The Table shows shows the OLS estimates of β from Equation (1) for interventions by non-political influencers in the debate. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

lowers than politicians. For example, the 60th member in this list by number of followers (singer songwriter Shawn Mendes) has more than 26 millions followers. The analysis on this set of such prominent influencers therefore may provide an upper bound estimate of the effect due to news amplification. Many of the included public figures, such as Kanye West, Kim Kardashian, or Ellen Degeneres, are controversial and divisive figures, perhaps even more so than some of the politicians included in our list. The estimates should thus also capture effects due to divisiveness.

We estimate the event-study specification (Equation (3)) and Equation (1) by considering p_e as the timing of the first intervention by any of these non-political public figures. The overall net effects in the post-intervention period are undistinguishable from zero, both for partisanship and tribalism, as shown in Table 4. The event-study results, displayed in the left panel of Appendix Figure A.11, show that the levels of partisanship and tribalism in the debate remain stable after a non-political intervention. For tribalism, the estimates show a small and short-lived increase in tribalism, which occurs two hours after the intervention. However, this is due to a single event, and is not robust to the exclusion of this outlier from the analysis set, as shown in the right panel of Appendix Figure A.11. Overall, the results suggest that interventions by non-political public figures have no consistent influence on the partisan or tribal nature of the public debate.

6.2 Political supply of partisan rhetoric polarizes public debates

We now show that politicians polarize the public debate partly because they supply a rhetoric that casts news events through a partisan lens. We proceed in two steps.

First, we characterize how the rhetoric of political interventions differs from that of other influencers, traditional news media organizations, as well as regular social media users, precisely in that they supply partisan worldviews. Column 1 of Table B.6 shows that the first political intervention by a politician is 18 to 20 percentage points more likely to contain partisan language compared to the first intervention by a traditional news media organization or by a non-politician influencer, or compared with a tweet by a regular user prior to a political intervention. Results in Column 5 that consider all (not just first) interventions by politicians and non-political influencers confirm this pattern. Columns 2 and 6 show that interventions by politicians are also more likely to contain tribal language compared to a pre-treatment tweet by a regular user (by 36 to 44 percentage points), but the difference between politicians and non-politicians influencers in the use of tribal language is insignificant (as indicated at the bottom of the table). However, non-politicians influencers only use tribal rhetoric when, and after, politicians do so: the probability that a non-politician influencer uses tribal rhetoric is 7% and statistically indistinguishable from zero when the first political intervention does not use tribal rhetoric. By contrast, when the first political intervention uses tribal rhetoric, 52% of interventions by non-politicians influencers also use tribal rhetoric.

Second, we show that partisan political interventions have a disproportionately polarizing effect. To show this, we modify Equation (1) to estimate the polarizing effects of interventions with and without partisan or tribal rhetoric. We estimate:

$$y_{i,e,t} = \beta_S S_{i,e,t}^{(p_e,T)} + \beta_{NS} NS_{i,e,t}^{(p_e,T)} + \eta_e + \theta_t + \zeta t + \upsilon_{i,e,t} \tag{5}$$

where $S_{i,e,t}^{(N,M)}$ (respectively $NS_{i,e,t}^{(N,M)}$) takes value 1 between times N and M if the political intervention is of type S (respectively, not of type S), where S measures whether the tweet contains a given rhetoric. We systematically report statistical tests of the difference between coefficients β_S and β_{NS} to test how the supply of partisanship and tribalism shapes the public debate.

The first panel of Table 5 displays the estimation results. Column 1 shows that political interventions that supply partisanship are associated with a 1.56 percentage points increase in the debate's partisan content after the intervention. Political interventions that do not supply partisan language are associated with a more modest increase of $.68$ per-

Table 5: Politicians' rhetoric and characteristics and partisanship and tribalism of the public debate

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
SUPPLY=0 * POST-Intervention	0.682**	1.692	0.900**	2.107
	(0.277)	(1.431)	(0.374)	(1.466)
SUPPLY=1 * POST-Intervention	1.563***	4.741**	0.750**	2.155
	(0.490)	(2.111)	(0.348)	(1.887)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025
Supply	Partisanship	Partisanship	Tribalism	Tribalism
Difference	0.882	3.049	-0.150	0.049
SE	0.510	1.846	0.457	2.035
Type=0 * POST-Intervention	0.717**	1.097	0.748**	0.392
	(0.306)	(1.393)	(0.326)	(1.489)
Type=1 * POST-Intervention	1.188***	6.342***	0.829**	2.667*
	(0.264)	(1.465)	(0.320)	(1.534)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025
Type	Republican	Republican	Male	Male
Difference	0.471	5.245	0.081	2.274
SE	0.283	1.211	0.380	1.473

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (columns 1 and 3), or tribal language (columns 2 and 4). Panel A shows the results of β_S and β_{NS} from OLS estimation of Equation 5. In columns 1 and 2 of the top panel, S captures whether the political intervention contains partisan language. In columns 3 and 4 of the top panel, S captures whether the political intervention contains partisan language. In columns 1 and 2 of the bottom panel, S captures whether the political intervention is from a Republican. In columns 3 and 4 of the bottom panel, S captures whether the political intervention is from a male politician. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

centage points, a difference that is statistically significant at the 10% level—see bottom of panel. Column 2 shows that political interventions that supply partisan language are also associated with a large and statistically significant 4.74 percentage points increase in the *tribal* content of public debate. By contrast, as shown in Columns 3 and 4, political interventions that supply tribal language are not associated with any statistically significant change in either the partisan or tribal content of the debate compared to interventions that do not include partisan language. Overall, this evidence suggests that the supply of partisan rhetoric by politicians contributes to online polarization.

We investigate the role of politicians' characteristics in the bottom panel of Table 5. Partisan or gender differences among politicians have no bearing on the partisanship of

the ensuing debate (Columns 1 and 3), but Columns 2 and 4 suggest that debates in which the first intervention is by a Republican or by a male politician become more tribal.[30,31]

6.3 Who is polarized by political interventions?

Finally, we investigate which users are most involved in the online polarization of the debate after a political intervention. First, we study whether the polarizing effect of political interventions affects primarily already politicized users who are mobilized by the intervention, or non-politicized users who are infected by the political supply of partisan rhetoric. Second, we study whether the polarizing effect of political interventions primarily affects the content supplied by users already engaged in the debate prior to the intervention (perhaps highly informed users who follow news events more closely) or instead it affects mostly the content supplied by *newcomers* to the debate: users who first tweet about the event after the intervention. To do so, we estimate Equation (5) where $S_{i,e,t}^{(N,M)}$ takes value 1 between times N and M if the tweet originates from a "political user", which we define as a user who uses political identity terms in their profile description.[32] To distinguish between users already engaged in the debate and newcomers, we also estimate a version of Equation (5) where $S_{i,e,t}^{(N,M)}$ takes value 1 if the tweet originates from a user who had already tweeted about the debate before the political intervention, and $NS_{i,e,t}^{(N,M)}$ takes value 1 between times N and M if the tweet is from a newcomer to the debate. Estimation results, displayed in Table B.10 show that the polarizing effect of political interventions is *not* different across political and non-political users, and mostly driven by newcomers to a debate. One possible interpretation of this result is that political interventions "attract" users to the debate, or that they disproportionately attract more politicized and polarized users. However, we do not find that political interventions are associated with an increase in the volume of tweets, nor to an increase in the share of tweets by political users (Figure A.12). Therefore, our results are consistent with the idea that the supply of partisan rhetoric by politicians places mass-shooting events under a polarized lens that steers the view and polarizes the behavior of users who join the debate only after the intervention.

A last set of results, displayed in Table B.9, suggests amplified and accelerated effects

[30] Estimates for Equation (2) are consistent and shown in Table B.7 and Table B.8.

[31] Table B.6 shows that Republican and male politicians are less likely to supply partisan rhetoric compared with Democrat and female politicians but do not differ in the tribal rhetoric of their tweets.

[32] The list of political identity terms consists of terms such as Democrat, Republican, MAGA, socialist. See Appendix C.0.4

when the event takes place outside working hours and on weekends.[33]

7 Conclusion

One of democracy's objectives and virtues, as famously discussed by Madison (The Federalist Papers, 10), is to limit unhinged political polarization—the "violence of faction"—by funneling ideological conflicts through institutionalized peaceful debates. Today as in the past, rising political polarization poses a threat to democratic institutions. This paper documents that the polarization of the American political debate is partly the result of a top-down mechanism, whereby political leaders (whether intentionally or accidentally) poison the public debate. We show this effect within the context of political interventions on online debates regarding an important class of salient policy-relevant events: mass shootings. We show that politicians offer a partisan lens through which to look at these events and citizens who join the debate *after* the politicians are more likely to discuss the events through these partisan lens and more likely to use tribal language.

Many commentators have speculated that mass political polarization in America is the result of (or even engineered by) choices made by political leaders and our results are consistent with the view that elite polarization precedes mass polarization (e.g., Bueno de Mesquita and Dziuda, 2023; Callander and Carbajal, 2022; Canen et al., 2021; Handan-Nader et al., 2024; Phillips et al., 2024). While our data cannot shed light on the extent to which political leaders intentionally generate political polarization, we provide direct evidence that their communication efforts cause greater polarization. We speculate that our results are particularly evident in the context of social media precisely because social media facilitate and accelerate top-down mechanisms of polarization, by enabling direct and frequent contact between politicians and citizens. Thus, through this channel, the rise of social media may have naturally coincided with an increase in mass polarization. Normatively, our results also refocus attention from bottom-up causes of polarization (echo-chambers, misinformation, or the flattening of the informational environment) to the role played by elites, and in particular political elites.

References

Abramowitz, A. I. and S. Webster (2016). The rise of negative partisanship and the nationalization of u.s. elections in the 21st century. *Electoral Studies 41*, 12–22.

[33]We use the time at the location of the event to define non-work hours. Non-working hours are defined as hours of the day between 6 p.m. and 8 a.m. 26 out of 57 events occur during non-working hours. We define Saturday and Sunday as the weekend. 17 out of 57 events occur during the weekend.

Acemoglu, D., A. Ozdaglar, and A. ParandehGheibi (2010). Spread of (mis)information in social networks. *Games and Economic Behavior 70*(2), 194–227.

Adena, M., R. Enikolopov, M. Petrova, V. Santarosa, and E. Zhuravskaya (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics 130*(4), 1885–1939.

Alatas, V., A. G. Chandrasekhar, M. Mobius, B. A. Olken, and C. Paladines (2019). When celebrities speak: A nationwide twitter experiment promoting vaccination in indonesia. Technical report, *National Bureau of Economic Research*.

Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives 31*(2), 211–236.

Amarasinghe, A. and P. A. Raschky (2022). Competing for attention–the effect of talk radio on elections and political polarization in the us. *arXiv preprint arXiv:2206.13675*.

Angelucci, C., J. Cagé, and M. Sinkinson (2024). Media competition and news diets. *American Economic Journal: Microeconomics 16*(2), 62–102.

Angelucci, C. and A. Prat (2024). Is journalistic truth dead? measuring how informed voters are about political news. *American Economic Review 114*(4), 887–925.

Anger, I. and C. Kittl (2011). Measuring influence on twitter. In *Proceedings of the 11th international conference on knowledge management and knowledge technologies*, pp. 1–4.

Ash, E., S. Galletta, M. Pinna, and C. Warshaw (2024). From viewers to voters: Tracing fox news' impact on american democracy. *Journal of Public Economics 240*.

Bakshy, E., J. M. Hofman, W. A. Mason, and D. J. Watts (2011). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65–74.

Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to ideologically diverse news and opinion on facebook. *Science 348*(6239), 1130–1132.

Barber, M. and N. McCarty (2015). Causes and consequences of polarization. In *Solutions to Political Polarization in America*. Cambridge: Cambridge University Press.

Barilari, F. (2024). Shooting political polarization.

Beknazar-Yuzbashev, G., R. Jiménez Durán, J. McCrosky, and M. Stalinski (2022). Toxic content and user engagement on social media: Evidence from a field experiment. *Available at SSRN 4307346*.

Boffa, F., G. Battiston, E. Levi, and S. Stillman (2024). Strategic use of social media by political parties: Evidence from italy. Technical report, Free University of Bozen/Bolzano.

Borusyak, K., X. Jaravel, and J. Spiess (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, rdae007.

Bueno de Mesquita, E. and W. Dziuda (2023). Partisan traps. Technical report, National Bureau of Economic Research.

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019). Social media and xenophobia: evidence from russia. Technical report, *National Bureau of Economic Research*.

Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review 110*(11), 3522–48.

Cagé, J., N. Hervé, and B. Mazoyer (2022). Social media and newsroom production decisions.

Callander, S. and J. C. Carbajal (2022). Cause and effect in political polarization: A dynamic analysis. *Journal of Political Economy 130*(4), 825–880.

Campante, F., R. Durante, and A. Tesei (Eds.) (2023). *The Political Economy of Social Media*. CEPR E-book. CEPR Press.

Canen, N. J., C. Kendall, and F. Trebbi (2021). Political parties as drivers of us polarization: 1927-2018. Technical report, National Bureau of Economic Research.

Couttenier, M., S. Hatte, M. Thoenig, and S. Vlachos (2024). Anti-muslim voting and media coverage of immigrant crimes. *Review of Economics and Statistics 106*(2), 576–585.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014). Cross-border media and nationalism: Evidence from serbian radio in croatia. *American Economic Journal: Applied Economics 6*(3), 103–132.

DellaVigna, S. and E. Kaplan (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics 122*(3), 1187–1234.

Demszky, D., N. Garg, R. Voigt, J. Zou, J. Shapiro, M. Gentzkow, and D. Jurafsky (2019, June). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. pp. 2970–3005.

Ederer, F., P. Goldsmith-Pinkham, and K. Jensen (2023). Anonymity and identity online. *Working paper. Available at* `https://florianederer.github.io/ejmr.pdf`.

Enikolopov, R., M. Petrova, G. Russo, and D. Yanagizawa-Drott (2024). Socializing alone: How online homophily has undermined social cohesion in the us. *Available at SSRN*.

Enke, B. (2020). Moral values and voting. *Journal of Political Economy 128*(10), 3679–3729.

Finkel, E. J., C. A. Bail, M. Cikara, P. H. Ditto, S. Iyengar, S. Klar, L. Mason, M. C. McGrath, B. Nyhan, D. G. Rand, et al. (2020). Political sectarianism in america. *Science 370*(6516), 533–536.

Fowler, A., S. J. Hill, J. B. Lewis, C. Tausanovitch, L. Vavreck, and C. Warshaw (2022). Moderates. *American Political Science Review*, 1–18.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–74.

Gentzkow, M. and J. M. Shapiro (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics 126*(4), 1799–1839.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica 87*(4), 1307–1340.

Giavazzi, F., F. Iglhaut, G. Lemoli, and G. Rubera (2024). Terrorist attacks, cultural incidents, and the vote for radical parties: Analyzing text from twitter. *American Journal of Political Science 68*(3), 1002–1021.

Gillani, N., A. Yuan, M. Saveski, S. Vosoughi, and D. Roy (2018). Me, my echo chamber, and i: introspection on social media polarization. In *Proceedings of the 2018 World Wide Web Conference*, pp. 823–831.

González-Bailón, S., D. Lazer, P. Barberá, M. Zhang, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Freelon, M. Gentzkow, A. M. Guess, et al. (2023). Asymmetric ideological segregation in exposure to political news on facebook. *Science 381*(6656), 392–398.

Graham, J., J. Haidt, and B. A. Nosek (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology 96*(5), 1029.

Grosjean, P., F. Masera, and H. Yousaf (2023). Inflammatory political campaigns and racial bias in policing. *The Quarterly Journal of Economics 138*(1), 413–463.

Guess, A. M., N. Malhotra, J. Pan, P. Barberá, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, D. Freelon, M. Gentzkow, et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science 381*(6656), 398–404.

Guess, A. M., B. Nyhan, and J. Reifler (2020). Exposure to untrustworthy websites in the 2016 us election. *Nature human behaviour 4*(5), 472–480.

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023). Curtailing false news, amplifying truth. *Mimeo PSE*.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2023). Political implications of the rise of mobile broadband internet. In F. Campante, R. Durante, and A. Tesei (Eds.), *The Political Economy of Social Media*. Paris & London: CEPR Press.

Haidt, J. and J. Graham (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research 20*(1), 98–116.

Handan-Nader, C., A. C. Myers, and A. B. Hall (2024). Polarization and state legislative elections. *American Journal of Political Science*.

Hatte, S., E. Madinier, and E. Zhuravskaya (2021). Reading twitter in the newsroom: How social media affects traditional-media reporting of conflicts.

Henry, E., E. Zhuravskaya, and S. Guriev (2022, August). Checking and sharing alt-facts. *American Economic Journal: Economic Policy 14*(3), 55–86.

Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science 22*, 129–146.

Jackson, M. O., S. Malladi, and D. McAdams (2022). Learning through the grapevine and the impact of the breadth and depth of social networks. *Proceedings of the National Academy of Sciences 119*(34), e2205549119.

Jacob, M. S., B. E. Lee, and G. Gratton (2024). From gridlock to polarization. *New Working Paper Series*.

Koleva, S. P., J. Graham, R. Iyer, P. H. Ditto, and J. Haidt (2012). Tracing the threads: How five moral concerns (especially purity) help explain culture war attitudes. *Journal of research in personality 46*(2), 184–194.

Lenz, G. S. and C. Lawson (2011). Looking the part: Television leads less informed citizens to vote based on candidates' appearance. *American Journal of Political Science 55*(3), 574–589.

Levitsky, S. and D. Ziblatt (2019). *How democracies die*. Crown.

Levy, G. and R. Razin (2019). Echo chambers and their effects on economic and political outcomes. *Annual Review of Economics 11*(1), 303–328.

Levy, R. (2021, March). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review 111*(3), 831–70.

Manacorda, M., G. Tabellini, and A. Tesei (2022). Mobile internet and the rise of political tribalism in europe.

Manacorda, M., G. Tabellini, and A. Tesei (2023). Mobile internet and the rise of communitarian politics. In F. Campante, R. Durante, and A. Tesei (Eds.), *The Political Economy of Social Media*. Paris & London: CEPR Press.

Martin, G. J. and A. Yurukoglu (2017). Bias in cable news: Persuasion and polarization. *American Economic Review 107*(9), 2565–2599.

Mattozzi, A., S. Nocito, and F. Sobbrio (2023). Fact-checking politicians. *Available at SSRN 4258130*.

McCarty, N., K. T. Poole, and H. Rosenthal (2016). *Polarized America: The dance of ideology and unequal riches*. mit Press.

Moskowitz, D. J., J. C. Rogowski, and J. M. S. Jr. (2024). Parsing party polarization in congress. *Quarterly Journal of Political Science 19*(4), 357–385.

Müller, K. and C. Schwarz (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association 19*(4), 2131–2167.

Müller, K. and C. Schwarz (2023). From hashtag to hate crime: Twitter and antiminority sentiment. *American Economic Journal: Applied Economics 15*(3), 270–312.

Nyhan, B., J. Settle, E. Thorson, M. Wojcieszak, P. Barberá, A. Y. Chen, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, et al. (2023). Like-minded sources on facebook are prevalent but not polarizing. *Nature 620*(7972), 137–144.

Phillips, C. H., J. M. Snyder Jr, A. B. Hall, et al. (2024). Who runs for congress? a study of state legislators and congressional polarization. *Quarterly Journal of Political Science*.

Rambachan, A. and J. Roth (2023). A more credible approach to parallel trends. *Review of Economic Studies 90*(5), 2555–2591.

Ridhwan, K. M. and C. A. Hargreaves (2021). Leveraging twitter data to understand public sentiment for the covid-19 outbreak in singapore. *International Journal of Information Management Data Insights 1*(2), 100021.

Rogers, N. and J. J. Jones (2021). Using twitter bios to measure changes in self-identity: Are americans defining themselves more politically over time? *Journal of Social Computing 2*(1), 1–13.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights 4*(3), 305–22.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics 225*(2), 175–199.

Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. Oxford University Press.

Thorp, H. H. and V. Vinson (2024). Context matters in social media. *Science 385*(6716), 1393–1393.

Wang, T. (2021a). Media, pulpit, and populist persuasion: Evidence from father coughlin. *American Economic Review 111*(9), 3064–3092.

Wang, T. (2021b). Waves of empowerment: Black radio and the civil rights movement. Technical report, Working paper.

Wang, T. (2023). The electric telegraph, news coverage and political participation. Technical report, National Bureau of Economic Research.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the rwandan genocide. *The Quarterly Journal of Economics 129*(4), 1947–1994.

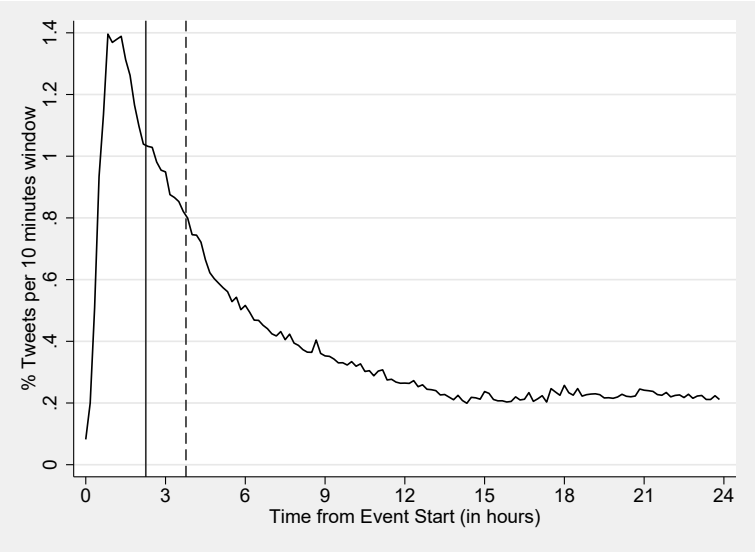
Yousaf, H. (2021). Sticking to one's guns: Mass shootings and the political economy of gun control in the united states. *Journal of the European Economic Association 19*(5), 2765–2802.

Yousefinaghani, S., R. Dara, S. Mubareka, A. Papadopoulos, and S. Sharif (2021). An analysis of covid-19 vaccine sentiments and opinions on twitter. *International Journal of Infectious Diseases 108*, 256–262.

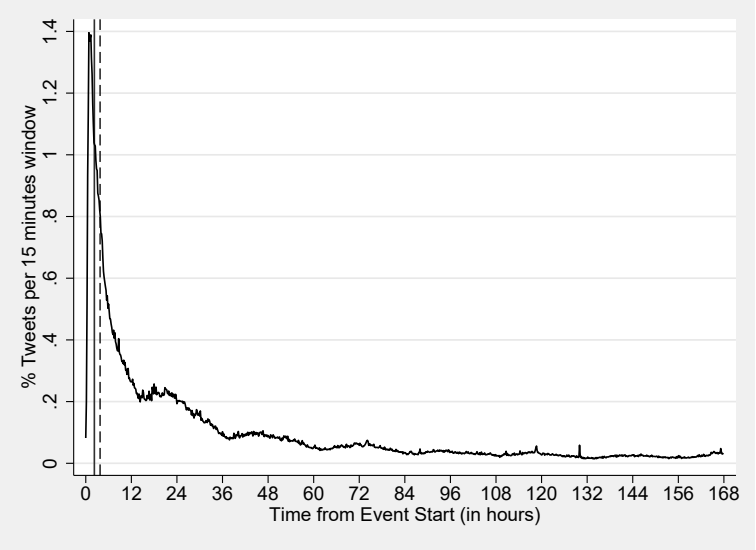
Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics 12*, 415–438.

Appendix

A Additional Figures



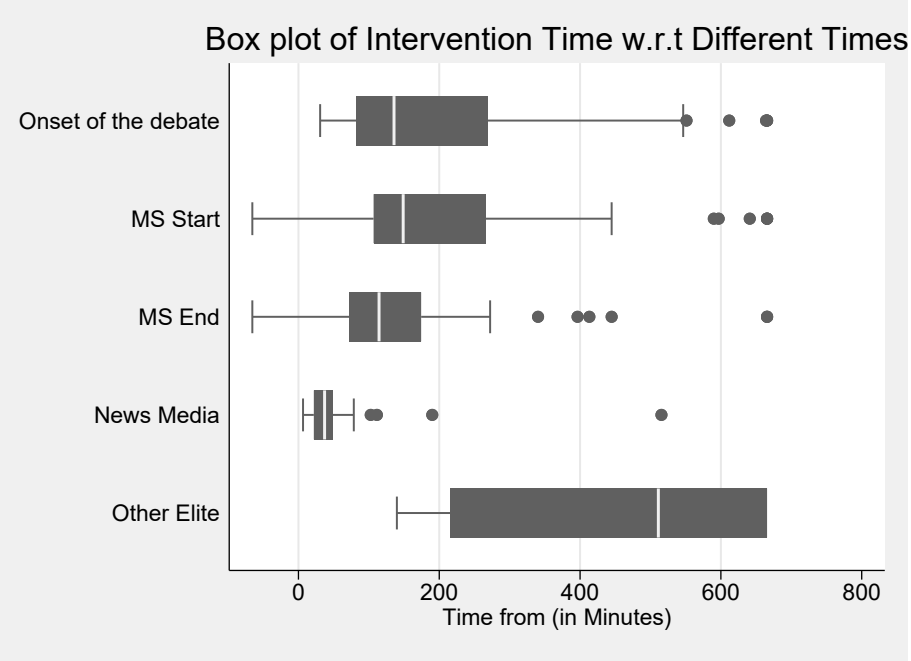
(a) In the first 24 hours



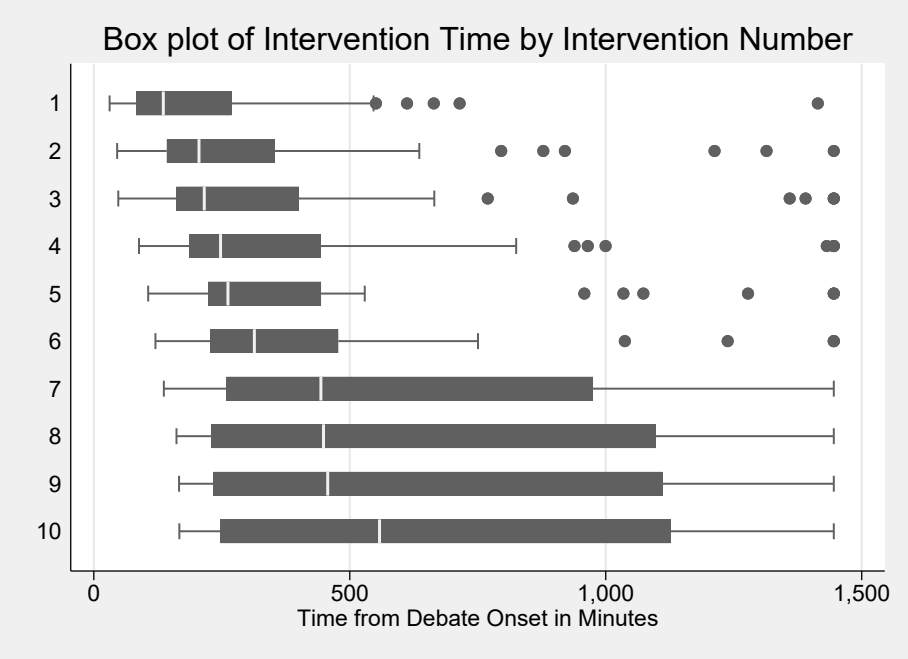
(b) In the entire sample

Figure A.1: Distribution of Tweets and the Timing of Political Interventions

These figures plot the distribution of tweets in each 10 minute bin from the onset of the debate. The y-axis plots the percentage of tweets that occur in each 10 minute bin. The x-axis plots the time from the onset of the debate (in hours). In panel (a) the figure zooms in the first 24 hours from the onset of the debate. In panel (b) the figure shows the distribution of tweets for the entire debate. In each panel, the solid vertical line indicates the median intervention time by a politician, while the dashed line indicates the average intervention time by a politician.



(a) Timing of First Intervention



(b) Timing of Subsequent Interventions

Figure A.2: Timing of Political Interventions

These figures plot the box-plot of timing of political interventions. In panel (a), we plot the 25th percentile, median, 75th percentile, and 1.5 times the inter-quartile range (boxplot) for the time of the first intervention w.r.t to five key times. In panel (b), we show the boxplot for the time of the first ten interventions w.r.t to start of the event. All times are winsorized to 95th percentile.

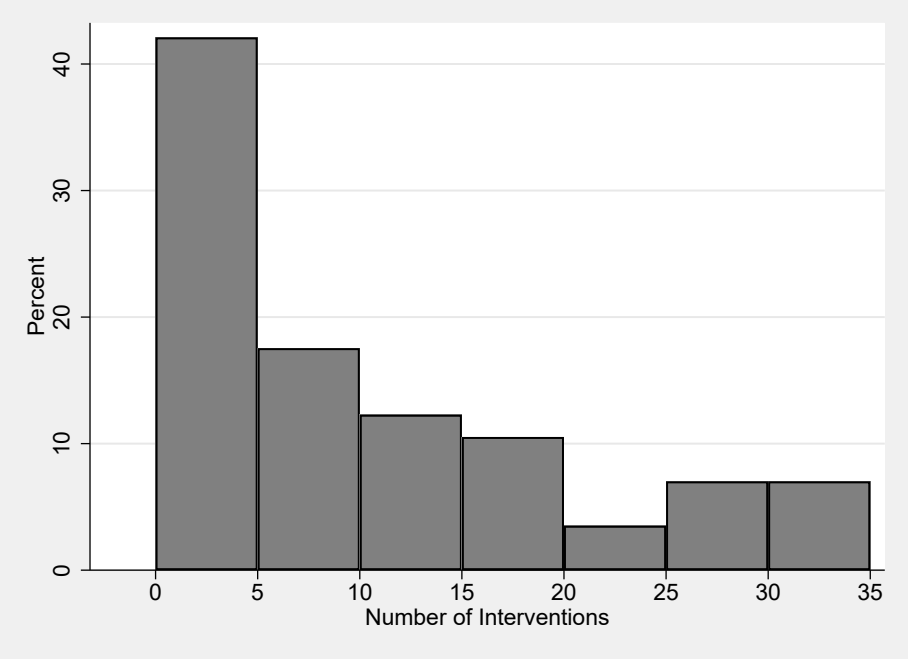
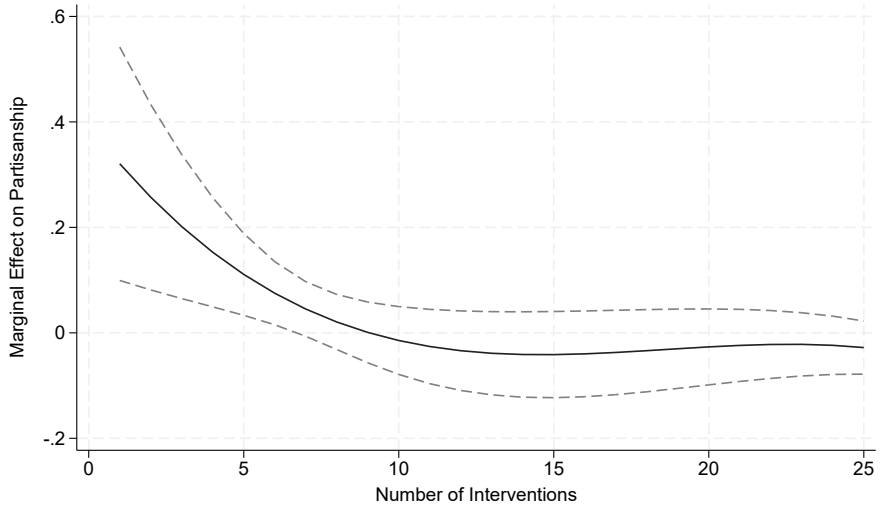
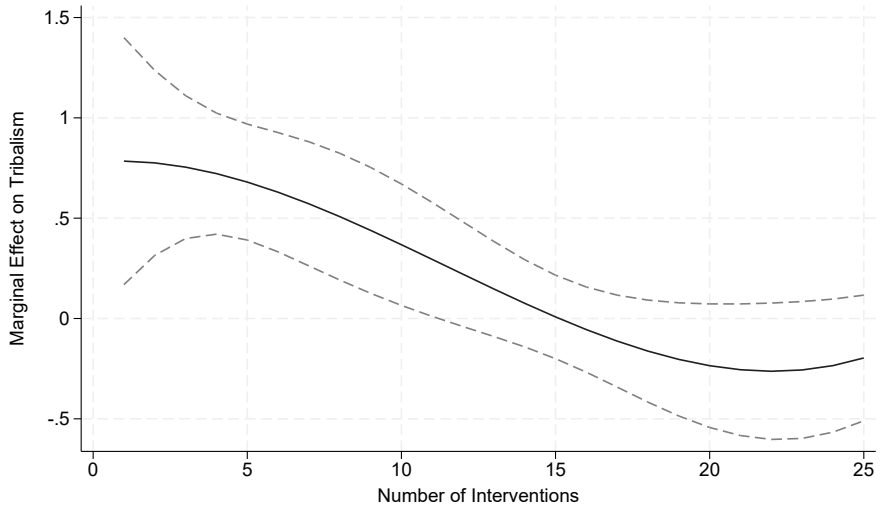


Figure A.3: Number of Political Interventions

The figure shows the distribution (histogram) of the number of unique political interventions in a public debate.



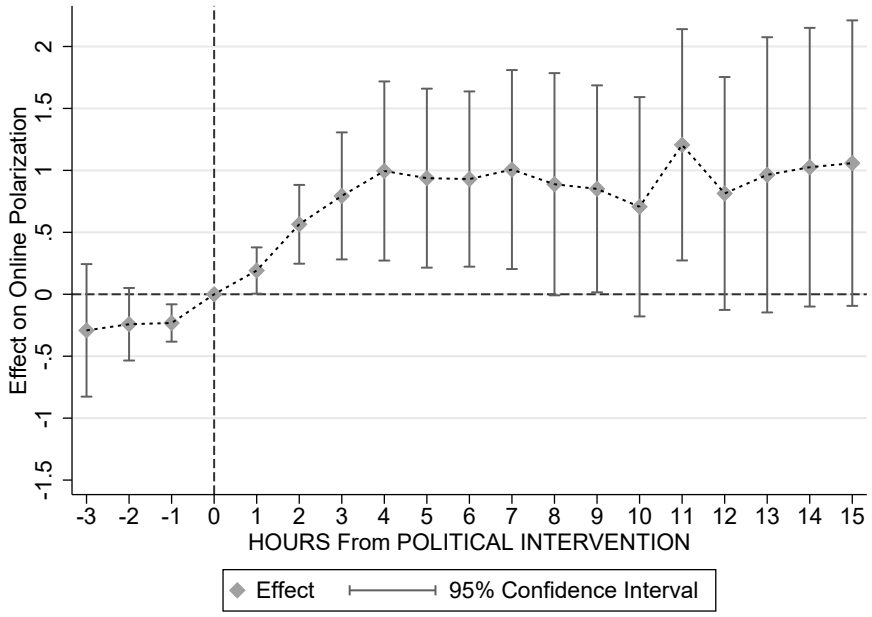
(a) Partisanship



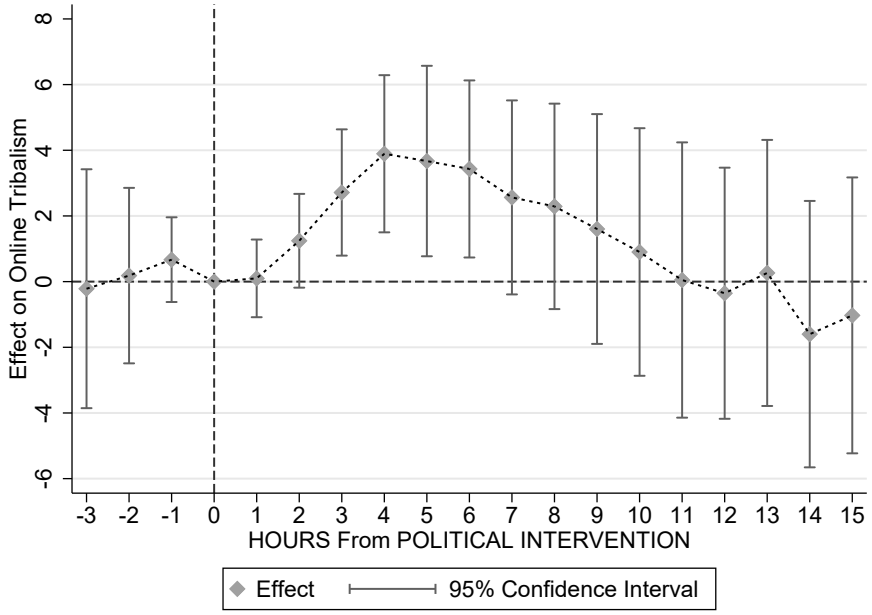
(b) Tribalism

Figure A.4: Number of political interventions: intensive margin results

These figures plot the marginal effect with 95% confidence intervals from OLS estimation of dependent variable as a quartic function of the number of political interventions. In panel (a), the dependent variable is Partisanship, in panel (b), the outcome variable is Tribalism. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate.



(a) Partisanship



(b) Tribalism

Figure A.5: Long-term event-study results

These figures plot OLS coefficients with 95% confidence intervals. The plotted coefficients are the event-studies coefficients associated with 1 hour windows, for a longer time horizon than our baseline specification. The outcome variable is Partisanship in panel (a) and Tribalism in panel (b). Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate.

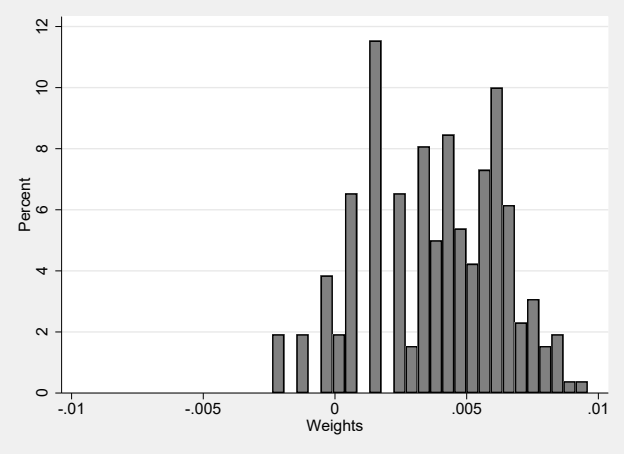
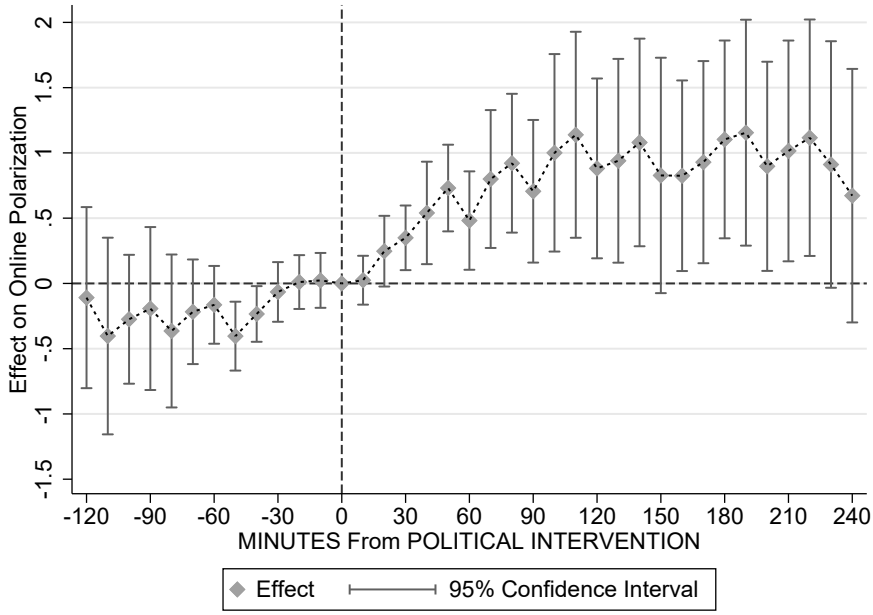
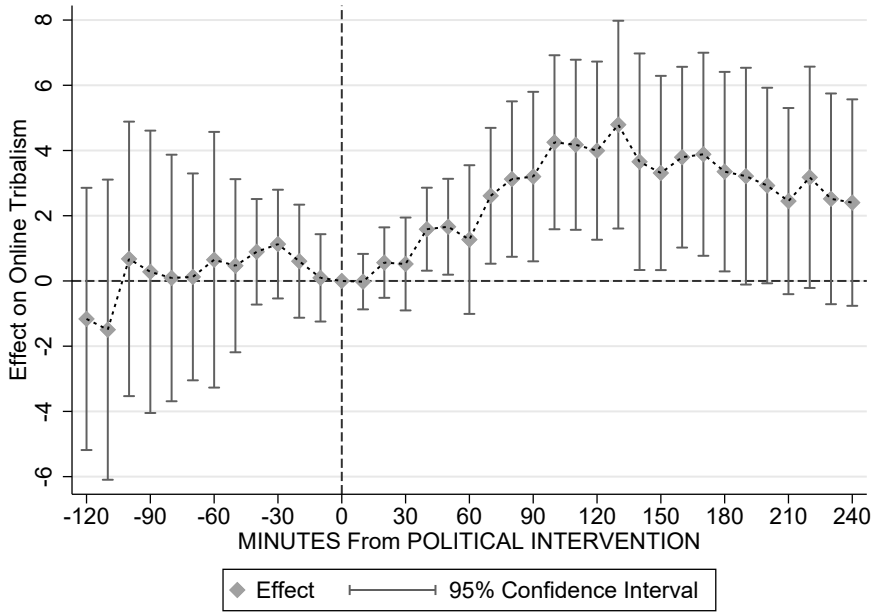


Figure A.6: DiD Weights (de Chaisemartin and D'Haultfoeuille (2020)

The figure shows the differences-in-differences weight computed using De Chaisemartin and d'Haultfoeuille (2020). The figure focuses on the short and medium-term impact of treatment. For calculating the weights, the data is converted into 30-minute intervals and restricted to 120 minutes from the onset of the treatment.



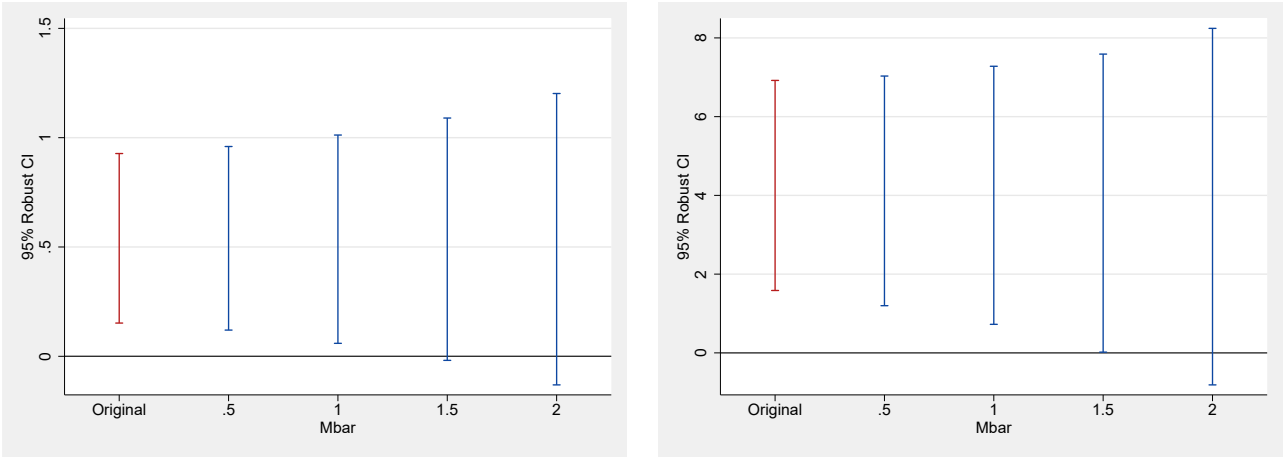
(a) Partisanship



(b) Tribalism

Figure A.7: Event-study results with no time trend

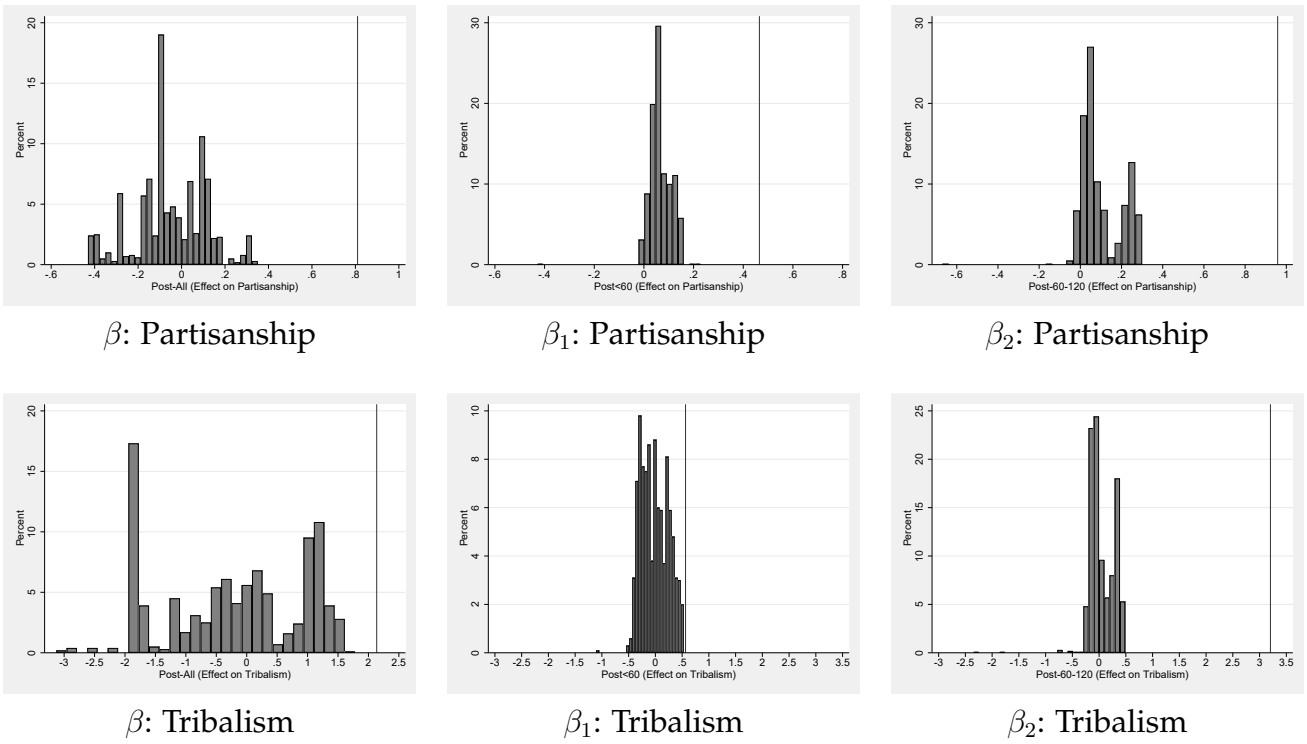
These figures plot OLS coefficients with 95% confidence intervals. The plotted coefficients are the β_τ coefficients associated with each 10 minute window from Equation (3), when we drop the time trend from Equation (3). The outcome variable is Partisanship in panel (a) and Tribalism in panel (b). Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate.



(a) Partisanship (b) Tribalism

Figure A.8: First political intervention, partisanship and tribalism: Robustness to parallel trends violations

The Figure displays robust confidence sets for β from Equation (1) for different values for the parameter $\bar{M}$ described in Rambachan and Roth (2023), which bound the maximal post-treatment violation of parallel trends by $\bar{M}$ times the maximal pre-treatment violation of parallel trends.

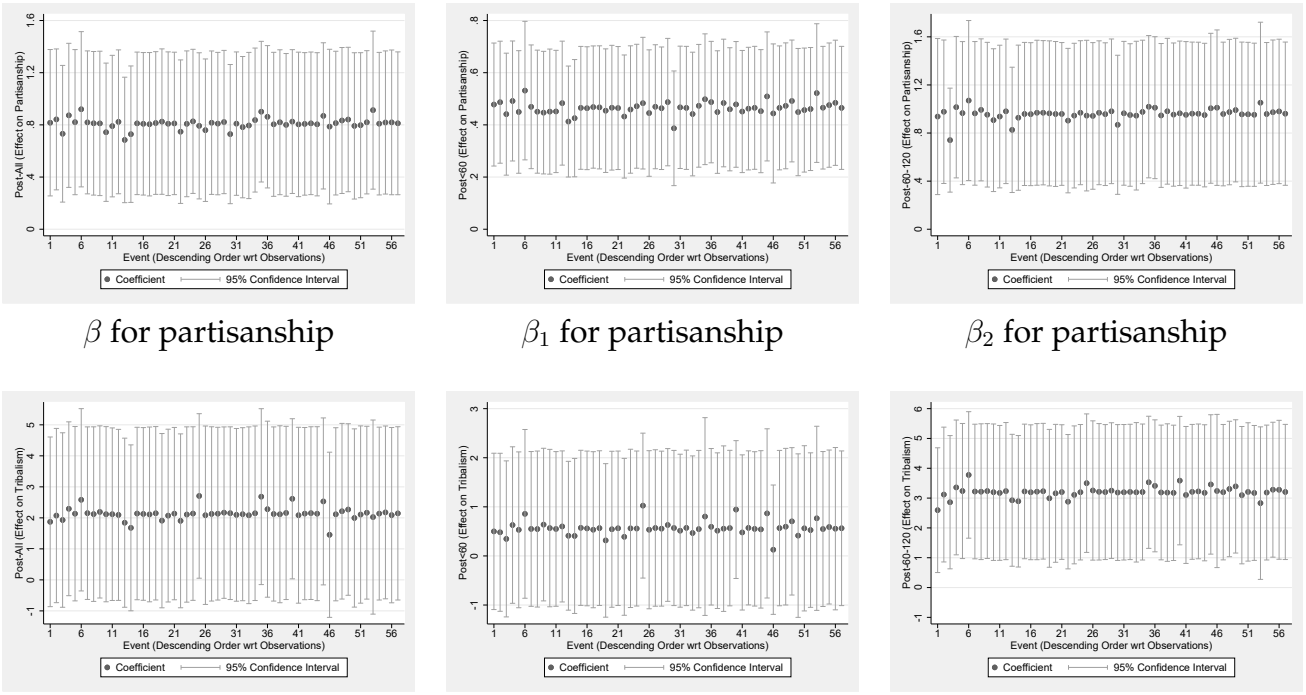


β: Partisanship | β_1: Partisanship | β_2: Partisanship

β: Tribalism | β_1: Tribalism | β_2: Tribalism

Figure A.9: First political intervention, partisanship and tribalism: Randomization inference

This figure shows the distribution of β (left panels), β_1 (middle panels), and β_2 (right panels) from Equations (1) and (2) for Partisanship (top panel) and tribalism (bottom panel) where instead of using the real distribution of political interventions, we randomly reallocate the same number of political interventions across events. The results of this permutation inference with placebo treatments are based on $1,000$ replications. The vertical bar indicates the coefficient obtained from the actual distribution of political interventions.

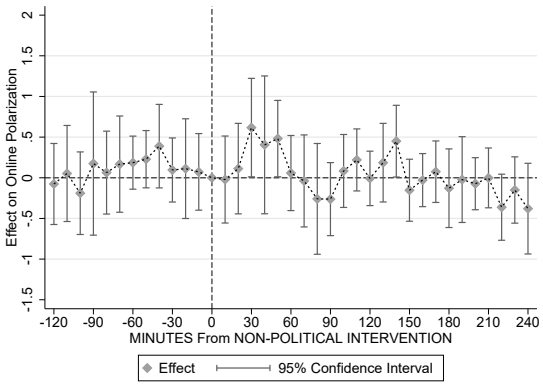


β for partisanship | β_1 for partisanship | β_2 for partisanship

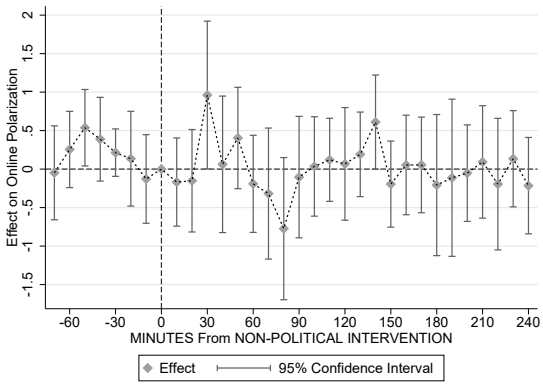
β for tribalism | β_1 for tribalism | β_2 for tribalism

Figure A.10: Estimates from Equation (2), dropping one event at a time

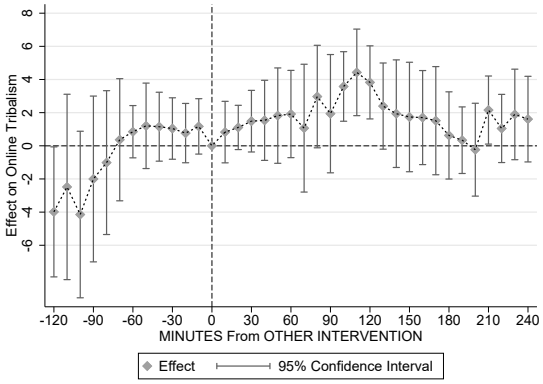
The Figure displays estimated β_1 and β_2 from different estimations of Equation (2), dropping one event at a time from the estimation sample.



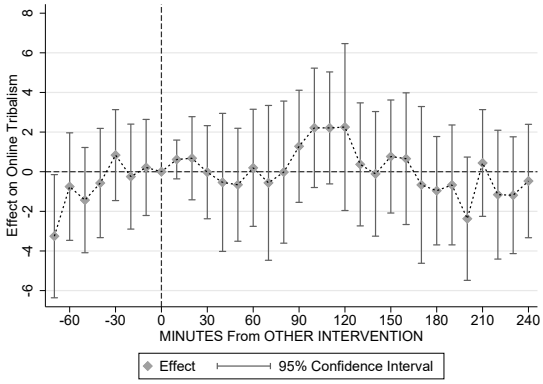
Partisanship



Partisanship, excluding one event



Tribalism



Tribalism, excluding one event

Figure A.11: Non-political interventions, partisanship and tribalism: Event-Study Results

These figures plot OLS coefficients with 95% confidence intervals. The plotted coefficients are the β_τ coefficients associated with each 10 minute window, described in Equation (3) when p_e indicates the time when a non-political elite tweets. The outcome variable is Partisanship in the top panel and Tribalism in the bottom panel. Left panel shows full sample estimates and right panel shows estimates when excluding one event from the estimation sample. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate.

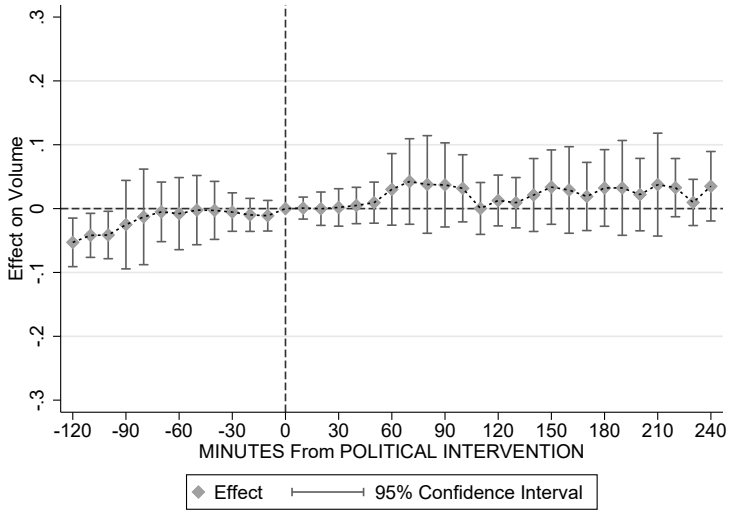


Figure A.12: Volume of Tweets for Political Users: Event-Study Results

These figures plot OLS coefficients with 95% confidence intervals. The plotted coefficients are the β_τ coefficients associated with each 10 minute window, described in Equation (3) when p_e indicates the time when a non-political elite tweets. The outcome variable is the number of tweets per 30-minute window. The sample size is limited to "political users" as defined in Section 6.3. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate.

B Additional Tables

Table B.1: Number of political interventions

Dependent Variable (DV): Number of Interventions			
	Mean of DV when:		
Independent Variable (X)	X=0	X=1	Difference
Panel A: Politician characteristics			
Male	6.88	13.06	6.17***
	(5.00)	(11.21)	(2.25)
Republican	8.59	18.38	9.79***
	(8.48)	(10.97)	(3.28)
SUPPLY: Partisan	11.74	8.1	-3.64
	(10.53)	(7.13)	(2.73)
SUPPLY: Tribal	10.83	11.3	0.48
	(9.92)	(10.35)	(2.83)
Followers Count: High	10.53	12.19	1.66
	(10.7)	(8.46)	(2.75)
Panel B: Event characteristics			
Shooter race: White	9.18	13.04	3.85
	(8.33)	(11.14)	(2.75)
Shooter race: Black	12.64	6.73	-5.91**
	(10.83)	(4.82)	(2.25)
Shooter race: Other	11.26	11.64	0.38
	(10.12)	(10.45)	(3.47)
Shooting location: School	11.54	10.17	-1.37
	(9.93)	(10.73)	(3.42)
Shooting location: Business	10.21	15.89	5.67
	(9.65)	(11.01)	(3.83)
Shooting location: Community	11.02	12.25	1.23
	(9.83)	(11.72)	(4.23)
Shooting deaths: High	6.54	20.29	13.75***
	(6.18)	(10.16)	(2.65)
Shooting length: Long	11.24	12.88	1.63
	(10.42)	(8.9)	(3.43)
Panel C: Twitter characteristics before intervention			
Tweets Volume: High	11.05	11	-0.05
	(9.99)	(10.41)	(3.12)
Posters' Followers: High	10.66	11.65	0.99
	(10.42)	(9.55)	(2.82)
Posters' Followings: High	9.79	13.21	3.42
	(9.42)	(10.89)	(2.98)
Event's Like Count: High	9.94	13.5	3.56
	(9.4)	(11.18)	(3.18)
Event's Retweet Count: High	10.42	11.95	1.53
	(9.33)	(11.12)	(2.94)

The unit of observation is an event. Column 1 (respectively, 2) shows the mean and standard deviation of the number of political interventions when the variable in the corresponding row is equal to 0 (respectively, 1). Column 3 displays the coefficients estimated from separate OLS regressions (with robust standard errors) of the number of first political interventions on each row variable. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.2: First political intervention, partisanship and tribalism: Robustness to excluding tweets within 1- or 2-min post intervention buffers

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
Equation 1				
POST-Intervention (β)	0.822***	2.152	0.831***	2.185
	(0.279)	(1.416)	(0.280)	(1.413)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025
Time Trend	Linear	Linear	Linear	Linear
Equation	1	1	1	1
Buffer in minutes	1	1	2	2
Equation 2				
POST-Intervention<1h (β_1)	0.479***	0.570	0.488***	0.602
	(0.121)	(0.802)	(0.123)	(0.803)
POST-Intervention 1-2h (β_2)	0.962***	3.204***	0.964***	3.212***
	(0.303)	(1.153)	(0.303)	(1.155)
POST-Intervention>2h (β_3)	0.986***	2.273	0.988***	2.280
	(0.361)	(1.504)	(0.361)	(1.504)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025
Time Trend	Linear	Linear	Linear	Linear
Equation	2	2	2	2
Mean DV	1.08	8.90	1.08	8.90
Buffer in minutes	1	1	2	2

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (Columns 1 and 3), or tribal language (Columns 2 and 4). Panel A shows the OLS estimates of β from Equation (1), but excluding tweets posted less than one or two minutes after the political intervention (as indicated). Panel B shows the OLS estimates of β_1, β_2, and β_3 from Equation (2) with $a = b = c = 60$, but excluding tweets posted less than one or two minutes after the political intervention (as indicated). Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. *** p<0.01, ** p<0.05, * p<0.1.

Table B.3: First political intervention, partisanship and tribalism: Robustness to event-specific time trends

VARIABLES	(1) Partisanship	(2) Tribalism
Equation 1		
POST-Intervention (β)	0.582**	2.301
	(0.264)	(1.496)
Observations	4,747,621	4,747,621
R-squared	0.015	0.027
Time Trend	Event $\times$ Linear	Event $\times$ Linear
Equation 2		
POST-Intervention<1h (β_1)	0.372***	0.615
	(0.118)	(0.775)
POST-Intervention 1-2h (β_2)	0.786**	3.153**
	(0.300)	(1.312)
POST-Intervention>2h (β_3)	0.718*	2.986
	(0.380)	(1.809)
Observations	4,747,621	4,747,621
R-squared	0.015	0.027
Time Trend	Event $\times$ Linear	Event $\times$ Linear
Mean DV	1.08	8.90

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (column 1), or tribal language (column 2). Panel A shows the OLS estimates of β from Equation (1), with the addition of event-specific time trends. Panel B shows the OLS estimates of β_1, β_2, and β_3 from Equation (2) with $a = b = c = 60$ and with the addition of event-specific time trends. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B.4: First political intervention, partisanship and tribalism: Robustness to alternative definitions of time fixed effects

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
Equation 1				
POST-Intervention (β)	0.809***	2.108	0.828***	2.306
	(0.285)	(1.428)	(0.264)	(1.399)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025
Time Trend	Linear	Linear	Linear	Linear
Equation	1	1	1	1
FE Window in minutes	15	15	60	60
Equation 2				
POST-Intervention<1h (β_1)	0.462***	0.514	0.484***	0.732
	(0.136)	(0.893)	(0.109)	(0.769)
POST-Intervention 1-2h (β_2)	0.960***	3.210***	0.953***	3.278***
	(0.317)	(1.190)	(0.279)	(1.099)
POST-Intervention>2h (β_3)	0.972**	2.215	1.000***	2.455*
	(0.377)	(1.556)	(0.330)	(1.458)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.026	0.014	0.025
Time Trend	Linear	Linear	Linear	Linear
Equation	2	2	2	2
Mean DV	1.08	8.90	1.08	8.90
FE Window in minutes	15	15	60	60

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (columns 1 and 3), or tribal language (columns 2 and 4). Panel A shows the OLS estimates of β from Equation (1) with θ_t defined as either 15- (columns 1 and 2) or 60-minutes (columns 3 and 4) time intervals since the onset of the debate (instead of 30 minutes as in our baseline specification). Panel B shows the OLS estimates of β_1, β_2, and β_3 from Equation (2) with $a = b = c = 60$ and with θ_t defined as either 15- or 60-minutes time intervals since the onset of the debate (instead of 30 minutes as in our baseline specification). Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B.5: First political intervention, partisanship and tribalism: Robustness to changing the set of politicians

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
Equation 1				
POST-Intervention (β)	0.904***	2.548*	0.896***	2.588*
	(0.292)	(1.380)	(0.300)	(1.341)
Observations	4,747,157	4,747,157	4,746,233	4,746,233
R-squared	0.014	0.025	0.014	0.025
Time Trend	Linear	Linear	Linear	Linear
Equation	1	1	1	1
Politicians with followers	>500k	>500k	>200k	>200k
Equation 2				
POST-Intervention<1h (β_1)	0.549***	0.904	0.265*	0.155
	(0.157)	(0.708)	(0.152)	(0.519)
POST-Intervention 1-2h (β_2)	1.119***	3.809***	0.744***	2.890***
	(0.321)	(1.045)	(0.185)	(0.580)
POST-Intervention>2h (β_3)	1.222***	3.028**	1.123***	2.524***
	(0.364)	(1.463)	(0.258)	(0.749)
Observations	4,746,233	4,746,233	4,746,233	4,746,233
R-squared	0.014	0.025	0.014	0.025
Time Trend	Linear	Linear	Linear	Linear
Equation	2	2	2	2
Mean DV	1.08	8.90	1.08	8.90
Politicians with followers	>500k	>500k	>200k	>200k

The unit of observation is a tweet. We now consider the first political intervention by any politician with at least 500,000 (columns 1 and 2) or 200,000 followers (column 3 and 4). The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (Columns 1 and 3), or tribal language (Columns 2 and 4). Panel A shows the OLS estimates of β from Equation (1). Panel B shows the OLS estimates of β_1, β_2, and β_3 from Equation (2) with $a = b = c = 60$. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B.6: Politicians supply partisanship, while other elites do not

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism	(5) Partisanship	(6) Tribalism	(7) Partisanship	(8) Tribalism
Political	17.686***	35.843***			18.840***	43.915***		
	(5.374)	(6.838)			(1.327)	(1.655)		
Non-Political	-1.182***	54.197***			2.267	49.181***		
	(0.019)	(14.504)			(3.388)	(9.146)		
News	-0.882***	-4.010***			1.384***	5.469***		
	(0.102)	(0.417)			(0.089)	(0.202)		
Republican			-19.535**	7.597			-5.153*	-2.413
			(7.634)	(17.534)			(2.792)	(3.478)
Male			-18.317	-15.099			-8.673**	3.670
			(13.656)	(15.985)			(4.353)	(4.792)
Observations	324,967	324,967	52	52	355,968	355,968	909	909
R-squared	0.001	0.001	0.122	0.019	0.008	0.008	0.013	0.001
Mean Supply	1.18	10.06	19.23	44.23	1.36	10.06	1.36	10.06
Intervention	First	First	First	First	All	All	All	All
POL-NOPOL	18.868	-18.353			16.574	-5.265		
SE	5.374	16.035			3.639	9.294		
POL-NEWS	18.568	39.853			17.456	38.447		
SE	5.375	6.850			1.330	1.666		
Mean Republican			0.67	0.67			0.75	0.75
Mean Male			0.25	0.25			0.20	0.20

The unit of observation is a tweet by a politician or by another elite. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (odd columns), or tribal language (even columns). In columns 1 and 2 (respectively 5 and 6), the sample consists of all tweets prior to the first political intervention as well as first (respectively all) interventions by any politician or non-political influencer. In columns 3 and 4 (respectively 7 and 8), the sample consists of first (respectively all) interventions by politicians. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.7: Politicians' rhetoric and partisanship and tribalism of the public debate

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
SUPPLY=0 * POST-Intervention<1h	0.423**	0.538	0.534*	-0.099
	(0.182)	(1.378)	(0.280)	(0.874)
SUPPLY=0 * POST-Intervention 1-2h	0.974***	3.449**	1.277**	2.448
	(0.353)	(1.368)	(0.542)	(1.676)
SUPPLY=0 * POST-Intervention>2h	0.957**	2.321	1.212**	3.654*
	(0.415)	(1.681)	(0.539)	(2.134)
SUPPLY=1 * POST-Intervention<1h	1.525***	4.631***	0.662**	2.555
	(0.545)	(1.189)	(0.286)	(1.952)
SUPPLY=1 * POST-Intervention 1-2h	1.582***	5.208**	0.939**	5.433***
	(0.482)	(2.448)	(0.370)	(1.541)
SUPPLY=1 * POST-Intervention>2h	1.832***	5.351**	1.022**	3.138
	(0.552)	(2.596)	(0.423)	(1.930)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025
Supply	Partisanship	Partisanship	Tribalism	Tribalism
Difference in <1h	1.102**	4.093***	0.128	2.654
	(0.537)	(1.524)	(0.385)	(1.850)
Difference in 1-2h	0.608	1.759	-0.338	2.986
	(0.460)	(2.337)	(0.594)	(2.037)
Difference in >2h	0.875	3.030	-0.190	-0.516
	(0.542)	(1.892)	(0.480)	(2.136)

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (columns 1 and 3), or tribal language (columns 2 and 4). Panel A shows the results of β_S and β_{NS} from OLS estimation of Equation 5 adapted to fit Equation (2). In columns 1 and 2, S captures whether the political intervention contains partisan language. In columns 3 and 4, S captures whether the political intervention contains partisan language. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. *** p<0.01, ** p<0.05, * p<0.1.

Table B.8: Politicians' characteristics and partisanship and tribalism of the public debate

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
Type=0 * POST-Intervention<1h	0.475*	0.628	0.928***	2.265***
	(0.238)	(0.996)	(0.284)	(0.691)
Type=0 * POST-Intervention 1-2h	0.979**	2.385	1.137***	3.496***
	(0.450)	(1.639)	(0.337)	(1.192)
Type=0 * POST-Intervention>2h	0.986**	1.653	0.959**	0.478
	(0.441)	(1.957)	(0.464)	(2.066)
Type=1 * POST-Intervention<1h)	0.957***	4.477**	0.459**	0.701
	(0.279)	(2.168)	(0.227)	(1.550)
Type=1 * POST-Intervention 1-2h	1.436***	8.489***	1.045**	3.782**
	(0.280)	(1.026)	(0.397)	(1.524)
Type=1 * POST-Intervention>2h	1.439***	6.914***	1.090**	3.188*
	(0.411)	(1.594)	(0.441)	(1.757)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.025
Type	Republican	Republican	Male	Male
Difference in <1h	0.482	3.849	-0.470	-1.565
	(0.366)	(2.281)	(0.317)	(1.336)
Difference in 1-2h	0.457	6.104***	-0.093	0.286
	(0.408)	(1.570)	(0.406)	(1.498)
Difference in >2h	0.454	5.262	0.131	2.709
	(0.307)	(1.224)	(0.409)	(1.553)

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (columns 1 and 3), or tribal language (columns 2 and 4). Panel A shows the results of β_S and β_{NS} from OLS estimation of Equation 5 adapted to fit Equation (2). In columns 1 and 2, S captures whether the political intervention is from a Republican. In columns 3 and 4, S captures whether the political intervention is from a male politician. Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. *** p<0.01, ** p<0.05, * p<0.1.

Table B.9: Event timing and partisanship and tribalism of the public debate

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
A * POST-Intervention<1h	0.234	-0.804	0.317**	0.180
	(0.150)	(0.825)	(0.147)	(0.970)
B * POST-Intervention<1h	0.490***	2.237*	1.078***	2.605***
	(0.180)	(1.200)	(0.279)	(0.608)
A * POST-Intervention 1-2h	0.887**	2.940*	0.583**	2.628*
	(0.388)	(1.735)	(0.255)	(1.427)
B * POST-Intervention 1-2h	0.566***	3.185***	2.043**	5.415***
	(0.192)	(1.184)	(0.846)	(1.897)
A * POST-Intervention>2h	0.390	2.116	0.689*	0.620
	(0.434)	(1.771)	(0.382)	(2.029)
B * POST-Intervention>2h	1.111***	2.352	1.788***	6.405***
	(0.341)	(1.637)	(0.596)	(2.256)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.026
Time Trend	Linear	Linear	Linear	Linear
Equation	1	1	1	1
A	Work Time	Work Time	Weekday	Weekday
B	Other Time	Other Time	Weekend	Weekend
B-A $< 1h$	0.256	3.041**	0.760**	2.425**
SE	(0.254)	(1.920)	(0.514)	(2.236)
B-A $1 - 2h$	-0.321	0.245	1.461*	2.787
SE	(0.357)	(1.843)	(0.841)	(1.148)
B-A $> 2h$	0.721**	0.237	1.099**	5.785***
SE	(0.352)	(1.353)	(0.298)	(2.149)

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (columns 1 and 3), or tribal language (columns 2 and 4). The table shows the results from OLS estimation of Equation (5) adapted to fit Equation (2). In columns 1 and 2, A (B) captures whether the event occurred during working hours (non-working hours). In columns 3 and 4, A (B) captures whether the event occurred during the weekday (weekend). Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. *** p<0.01, ** p<0.05, * p<0.1.

Table B.10: Diffusion vs Mobilization: TWFE results

VARIABLES	(1) Partisanship	(2) Tribalism	(3) Partisanship	(4) Tribalism
POST-Intervention (β) $\times$ (User Type=0)	0.772***	2.042	1.023***	3.857***
	(0.276)	(1.410)	(0.282)	(1.386)
POST-Intervention (β) $\times$ (User Type=1)	1.200***	3.091*	0.148	-3.217**
	(0.339)	(1.692)	(0.262)	(1.485)
Observations	4,747,621	4,747,621	4,747,621	4,747,621
R-squared	0.014	0.025	0.014	0.029
Time Trend	Linear	Linear	Linear	Linear
Equation	2	2	2	2
User Type	Political	Political	Returning	Returning
Mean DV	1.08	8.90	1.08	8.90

The unit of observation is a tweet. The dependent variable is equal to 100 (and zero otherwise) if the tweet contains partisan language (columns 1 and 3), or tribal language (columns 2 and 4). Panel A shows the results of β_S and β_{NS} from OLS estimation of Equation 5. In columns 1 and 2 of the top panel, S captures whether the user was a political user (user with political text in account description). In columns 3 and 4 of the top panel, S captures whether the user was a returning user (user who already posted before the political intervention). Standard errors are two-way clustered at the event level and at the level of 30 minutes time intervals since the onset of the debate. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

C Data Appendix

We closely follow the data collection and pre-processing methodology of Demszky et al. (2019) to build our database of online public debates.

Data collection. We use the python package *snscrape*,[34] to retrieve tweets. We not not include retweets and only include tweets in English. For each event, we retrieve all public tweets posted within seven days of an event that contain at least one Event Specific keyword in column 3 of Table C.1 and at least one Event Type Specific keyword in column 4 of Table C.1 from the following list: ”gun”, ”shoot”, ”kill”, ”attack”, ”massacre”, ”victim”, ”terror”, ”violence”, ”crime” along with any lemmas relevant to the location or type of event e.g. ”school” if the event happened in a school. We keep all events with more than 3000 tweets after cleaning.

[34] `https://github.com/JustAnotherArchivist/snscrape`. The package snscrape is a common alternative to the Twitter API for retrieving data from Twitter. See (Ridhwan and Hargreaves, 2021; Yousefinaghani et al., 2021).

Table C.1: Mass Shooting Event Characteristics

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
1	UCLA shooting	California, CA	1/06/16 16:49	1/06/16 17:00	school	Non-White	3
2	Orlando nightclub shooting	Orlando, FL	12/06/16 6:02	12/06/16 9:14	community other	Non-White	80
3	Townville Elementary School shooting	Townville, SC	28/09/16 17:44	28/09/16 18:00	school	White	2
4	Palm Springs shooting	Palm Springs, CA	8/10/16 20:13	9/10/16 7:50	residential	Non-White	2
5	Fort Lauderdale airport shooting	Broward County, FL	6/01/17 17:53	6/01/17 17:54	other	Non-White	5
6	North Park Elementary School shooting	San Bernardino, CA	10/04/17 17:27	10/04/17 17:27	school	Non-White	3
7	Congressional baseball shooting	Alexandria, VA	14/06/17 11:09	14/06/17 11:20	other	White	1
8	Bronx-Lebanon Hospital attack	Bronx, NY	30/06/17 18:45	30/06/17 19:55	community other	Non-White	2
9	Las Vegas shooting	Paradise, NV	2/10/17 5:05	2/10/17 5:15	community other	White	61
10	Marshall County High School shooting	Benton, KY	23/01/18 13:57	23/01/18 14:06	school	White	2
11	Parkland high school shooting	Parkland, FL	14/02/18 19:21	14/02/18 19:27	school	White	17
12	Great Mills high school shooting	Great Mills, MD	20/03/18 11:55	20/03/18 11:55	school	White	2
13	YouTube headquarters shooting	San Bruno, CA	3/04/18 19:46	3/04/18 19:48	other	Non-White	1
14	Nashville Waffle House shooting	Nashville, TN	22/04/18 9:25	23/04/18 19:07	retail	White	4
15	Santa Fe High School shooting	Santa Fe, TX	18/05/18 12:32	18/05/18 13:02	school	White	10
16	Art All Night shooting	Trenton, NJ	17/06/18 6:45		community other	Non-White	1

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
17	Capital Gazette shooting	Annapolis, MD	28/06/18 18:34	28/06/18 18:38	other	White	5
18	Jacksonville Landing shooting	Jacksonville, FL	26/08/18 17:30	26/08/18 17:36	community other	White	3
19	Aberdeen, Maryland shooting	Aberdeen, MD	20/09/18 13:06	20/09/18 13:09	other	Non-White	4
20	Pittsburgh synagogue shooting	Pittsburgh, PA	27/10/18 13:54	27/10/18 15:08	place of worship	White	11
21	Thousand Oaks shooting	Thousand Oaks, CA	8/11/18 7:18	8/11/18 7:38	retail	White	13
22	Mercy Hospital shooting	Chicago, IL	19/11/18 21:00	19/11/18 21:20	community other	Non-White	4
23	Sebring shooting	Sebring, FL	23/01/19 17:30	23/01/19 19:28	retail	White	5
24	Aurora, Illinois shooting	Aurora, IL	15/02/19 19:24	15/02/19 20:59	other	Non-White	6
25	Poway synagogue shooting	Poway, CA	27/04/19 18:23	27/04/19 18:33	place of worship	White	1
26	University of North Carolina at Charlotte shooting	Charlotte, NC	30/04/19 21:43	30/04/19 21:44	school	White	2
27	STEM School Highlands Ranch shooting	Douglas County, CO	7/05/19 19:53	7/05/19 20:07	school	White	1
28	Virginia Beach shooting	Virginia Beach, VA	31/05/19 20:08	31/05/19 20:44	other	Non-White	13
29	Gilroy Garlic Festival shooting (28/7/19, utc is 29/7)	Gilroy, CA	29/07/19 0:40	29/07/19 0:45	community other	White	4
30	El Paso shooting	El Paso, TX	3/08/19 16:39	3/08/19 16:45	retail	White	23
31	Dayton shooting	Dayton, OH	4/08/19 5:05	4/08/19 5:06	retail	White	10
32	Midland-Odessa shootings	Midland/Odessa, TX	31/08/19 20:17	31/08/19 21:20	other	White	8
33	Saugus High School shooting	Santa Clarita, CA	14/11/19 15:38	14/11/19 15:38	school	Non-White	3
34	Fresno shooting	Fresno, CA	18/11/19 2:00		residential	Non-White	4

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
35	Naval Air Station Pensacola shooting	Pensacola, FL	6/12/19 11:43	6/12/19 23:58	other	Non-White	4
36	Jersey City shooting	Jersey City, NJ	10/12/19 17:21	10/12/19 20:47	retail	Non-White	6
37	West Freeway Church of Christ shooting	White Settlement, TX	29/12/19 16:50	29/12/19 16:50	place of worship	White	3
38	NYPD officers shooting	Bronx, NY	10/02/20 0:30	10/02/20 12:00		Non-White	0
39	Milwaukee brewery shooting	Milwaukee, WI	26/02/20 20:08	26/02/20 20:08	other	Non-White	6
40	Medical Clinic Attack	Buffalo, MN	9/02/21 16:52	9/02/21 17:00	community other	White	1
41	Atlanta Spa Shootings	Atlanta, GA	16/03/21 20:50	17/03/21 0:30	retail	White	8
42	Boulder Supermarket Shooting	Boulder, CO	22/03/21 20:30	22/03/21 21:28	retail	Non-White	10
43	Indianapolis FedEx Shooting	Indianapolis, IN	16/04/21 3:00	16/04/21 3:04	other	White	9
44	San Jose VTA Rail Yard Shooting	San Jose, CA	26/05/21 13:33	26/05/21 13:43	other	White	10
45	Collierville Kroger Shooting	Colleyville, TX	23/09/21 18:30	23/09/21 18:34	retail	Non-White	2
46	Oxford High School Shooting	Oakland, MI	30/11/21 17:51	30/11/21 17:55	school	White	4
47	Dumas Car Show Shooting	Dumas, AR			community other		1
48	Downtown Sacramento Shooting	Sacramento, CA	3/04/22 9:01	3/04/22 9:01	other		6
49	NYC Subway attack	Sunset Park, NY	12/04/22 12:24	13/04/22 17:42	community other	Non-White	0
50	Buffalo Supermarket Shooting	Buffalo, NY	14/05/22 18:31	14/05/22 18:36	retail	White	10
51	Laguna Woods Church Shooting	Orange County, CA	15/05/22 20:26	15/05/22 20:26	place of worship	Non-White	1
52	Robb Elementary School shooting	Uvalde, TX	24/05/22 16:28	24/05/22 17:50	school	Non-White	22

Event #	Event Name	Location	Event Start Time	Event End Time	Location Type	Shooter Ethnicity	# Deaths
53	Saint Francis Hospital Shooting	Tulsa, OK	1/06/22 21:56	1/06/22 22:00	community other	Non-White	5
54	South Street Philadelphia Shooting	Philadelphia, PA	5/06/22 3:31		other	Non-White	3
55	Chattanooga Shooting	Chattanooga, TN	5/06/22 6:42		other		3
56	Highland Park Parade Shooting	Chicago, IL	4/07/22 15:14	4/07/22 23:03	other	White	7
57	Greenwood Park Mall shooting	Greenwood, IN	17/07/22 21:56	17/07/22 21:56	retail	White	4

This table contains mass shooting event level characteristics for each event in our dataset. It contains the event name, location, start and end date and time of the mass shooting event, location type, shooter ethnicity, and number of deaths.

Table C.2: Mass Shooting Event Retrieval and Cleaning

Event #	Total Tweets	Event Specific Keywords	Event Type Specific Keywords	Politician Int	Non-Political Elite Int
1	63643	ucla, mainak sarkar, mainaksarkar, william scott klug, williamscottklug, william klug, williamklug	MS*, engineering, sarkar	1	1
2	1086045	orlando, pulse nightclub, pulsenightclub, omar mateen, omarmateen	MS*, pulse, nightclub	1	1
3	6049	townville, jesse osborne, jesseosborne	MS*, elementary, school	1	0
4	15489	palm springs, palmsprings, john felix, johnfelix, jose gilbert vega, josegilbertvega, josevega, jose vega, lesley zerebny, lesleyzerebny	MS*, police, officer	1	0
5	56959	broward county, santiago-ruiz, fort lauderdale	MS*, airport, santiago	1	1
6	31277	north park elementary, northpark, san bernardino, sanbernardino	MS*, school	1	1
7	58101	alexandria, congressional baseball, james hodgkinson, jameshodgkinson, eugene simpson stadium	MS*, hodgkinson	1	0
8	15937	bronx-lebanon hospital, bronx, bronx lebanon, bronxlebanon	MS*, hospital	1	0
9	582748	nevada, las vegas, lasvegas, steven paddock, stevenpaddock, route 91 harvest	MS*, route 91, harvest, paddock, festival	1	1
10	6756	draffenville, marshall county, marshallcounty	MS*, school	1	0
11	181676	stoneman douglas, stonemandouglas, parkland, nikolas cruz, nikolascru, nikolas jacob cruz, nikolasjacobcruz	MS*, school, marjory, stoneman	1	1
12	13053	st. mary's county, st marys, st mary's, great mills, greatmills	MS*, school	1	0
13	40614	san bruno, sanbruno, youtube headquarters, youtube hq	MS*	1	0
14	18342	nashville, waffle house, antioch, travis jeffrey reinking, travisjeffreyreinking	MS*, waffle	0	0
15	63569	santa fe, santafe, dimitrios pagourtzis, dimitriospagourtzis	MS*, school	1	1
16	3642	trenton, art all night, artallnight, amir armstrong, amirarmstrong, davonewhite, davone white, tahaij wells, tahaijwells	MS*, festival	0	0
17	85307	annapolis, capital gazette, capitalgazette, jarrod ramos, jarrodramos	MS*, capital	1	0

Event #	Total Tweets	Event Specific Keywords	Event Type Specific Keywords	Political Int	Non-Political Elite Int
18	69254	jacksonville, david katz, davidkatz, elijah clayton, elijahclayton, taylor robertson, taylorrobertson, madden nfl 19, maddennfl19	MS*, madden, tourney, tournament, twitch, stream, game, qualifier	1	1
19	6761	aberdeen, snochia moseley, snochiamoseley, riteaid, rite aid	MS*	1	0
20	225582	pittsburgh, robert gregory bowers, robertgregorybowers, tree of life, treeoflife	MS*, synagogue	1	0
21	62561	thousand oaks, thousandoaks, ian david long, iandavidlong	MS*	1	1
22	8961	mercy hospital, mercyhospital, juan lope, juanlopez	MS*	0	0
23	5521	sebring, zephen allen xaver, zephenallenxaver, zephen xaver, zephenxaver	MS*, bank, suntrust	1	0
24	23217	aurora, gary montez martin, garymontezmartin, henry pratt company, henryprattcompany	MS*	1	0
25	22162	poway, john timothy earnest, johntimothyearnest, johnearnest, john earnest, lori gilbert-kaye, lorigilbert-kaye	MS*, synagogue	1	0
26	38671	charlotte, unc, trystan andrew terrell, trystanandrewterrell, riley howell, rileyhowell, ellis parlier, ellisparlier	MS*, school, unc	1	0
27	6795	highlands ranch, highlandsranch, alec mckinney, alecmckinney, devon erickson, devonerickson	MS*, school, academy	1	0
28	41724	virginia beach, virginiabeach, dewayne craddock, dewaynecraddock	MS*, courthouse	1	0
29	64498	gilroy, garlic festival, santino william legan, santinowilliamlegan	MS*, garlic, festival	1	0
30	331273	el paso, elpaso, patrick wood crusius, patrickwoodcrusius, patrickcrusius, patrick crusius	MS*, walmart	1	0
31	181275	dayton, connor betts, connorbetts, connor stephen betts, connorstephenbetts	MS*	1	0
32	34783	odessa, midland odessa, midlandodessa, seth ator, sethator, seth aaron ator, sethaaronator	MS*	1	0
33	26660	santa clarita, santaclarita, saugus high school, saugushighschool, nathaniel berhow, nathanielberhow	MS*, school, saugus	1	1
34	5238	fresno	MS*	1	0

Event #	Total Tweets	Event Specific Keywords	Event Type Specific Keywords	Political Int	Non-Political Elite Int
35	36362	pensacola, mohammed saeed alshamrani, mohammedsaeedalshamrani, mohammed alshamrani, mohammedalshamrani	MS*	1	0
36	27782	jersey city, jerseycity, david anderson, davidanderson, francine graham, francinegraham, joseph seals, josephseals	MS*	1	0
37	8471	white settlement, west freeway church, westfreewaychurch, whitesettlement, keith thomas kinnunen, keiththomaskinnunen, jack wilson, jackwilson	MS*, church	1	0
38	3856	bronx, robert williams, robertwilliams	MS*, cop, police, nypd	1	0
39	18255	milwaukee, molson coors, molsoncoors, anthony ferrill, anthonyferrill	MS*, miller, molson, campus	1	0
40	3918	allina health, buffalo, buffalo crossroads	MS*, allina, clinic	1	0
41	89233	cherokee, acworth, atlanta, gold spa	MS*	1	1
42	90857	boulder	MS*, supermarket, king soopers	1	0
43	16212	fedex facility, indiana	MS*, fedex	1	0
44	22544	santa clara, vta, valley transportation authority, san jose	MS*, vta, yard	1	0
45	11329	collierville, tennessee, kroger, memphis	MS*	0	0
46	27243	oxford high, oxfordhighschool, oxford high school	MS*	1	0
47	4720	arkansas, dumas	MS*	0	0
48	24593	sacramento	MS*	1	0
49	50610	sunset park, brooklyn	MS*, subway, train, station	1	0
50	278672	buffalo	MS*, supermarket	1	1
51	9326	laguna woods	MS*	1	0
52	666645	uvalde, robb elementary	MS*	1	1
53	32341	saint francis hospital, tulsa	MS*	1	0
54	15569	south street, philadelphia	MS*	1	0
55	3765	chattanooga	MS*	1	0
56	160194	highland park	MS*	1	0
57	11580	greenwood park, greenwood	MS*, mall	1	0

This table contains information about the total number of tweets in each event, the Event Specific Keywords and Event Type Specific Keywords used for cleaning, Political Int =1 if there is an intervention in the event by a politician of interest and Non-Political Elite Int = 1 if a non-political elite intervenes. MS* = gun, shoot, kill, attack, massacre, victim, terror, violence, crime. Keywords are chosen to reflect the methodology of Demszky et al. (2019).

C.0.1 List of Partisanship phrases identified by Gentzkow et al. (2019) between 2005-2016 Congressional Sessions

afford care, african american, al qaeda, american energi, american peopl, busi owner, care act, care bill, care plan, care reform, children health, civil war, clean air, climat chang, colleagu join, colleagu support, comprehens immigr, continent shelf, credit card, death tax, depart homeland, dog coalit, employ mandat, farm bill, fornia madam, general kagan, god pleas, govern spend, govern takeov, gun violenc, hate crime, homeland secur, human traffick, hurrican katrina, illeg immigr, immigr reform, insur compani, interest rate, job creation, job creator, men women, mental health, middl class, million american, minimum wage, muslim brotherhood, nation debt, nation guard, natur gas, nobid contract, oil compani, outer continent, plan parenthood, pleas bless, presid health, progress caucus, public health, puerto rico, radic islam, rais tax, recoveri act, religi freedom, reserv balanc, side aisl, stem cell, stimulus bill, student loan, tax break, tax increas, tax rate, tax relief, taxpay dollar, troop iraq, unemploy benefit, unemploy insur, vote right, war iraq, war terror

C.0.2 List of Tribalism words from the loyalty/betrayal dimension of Moral Foundation theory by Graham et al. (2009) as used by Enke (2020)

abandon, ally, apostasy, apostate, betray, cadre, cliqu, cohort, collectiv, communal, commune, communis, communit, comrad, deceiv, deserted, deserter, deserting, devot, disloyal, enem, familial, families, family, fellow, foreign, group, guild, homeland, immigra, imposter, individual, insider, jilt, joint, loyal, member, miscreant, nation, patriot, renegade, segregat, sequester, solidarity, spy, terroris, together, traitor, treacher, treason, unison, unite

C.0.3 Interventions

Table C.3: Relevant Political Elite Twitter Accounts

Twitter Username	Followers	Party	Gender	Role	State/ Territory	Account Type
alfranken	1.1 M	D	M	Senator	Minnesota	Personal
amyklobuchar	1.9 M	D	F	Senator	Minnesota	Personal
aoc	13 M	D	F	Representative	New York	Personal
ayannapressley	1.1 M	D	F	Representative	Massachusetts	Personal
barackobama	132 M	D	M			Personal
berniesanders	15.5 M	I	M	Senator	Vermont	Personal

betoorourke	2.4 M	D	M	Representative	Texas	Personal
chrismurphyct	1.1 M	D	M	Senator	Connecticut	Office
corybooker	4.9 M	D	M	Senator	New Jersey	Personal
dancrenshawtx	1.2 M	R	M	Representative	Texas	Personal
devinnunes	1.3 M	R	M	Representative	California	Official
ewarren	5.9 M	D	F	Senator	Massachusetts	Personal
gavinnewsom	2 M	D	M	Governor	California	Personal
gopleader	1.5 M	R	M			
govrondesantis	2.5 M	R	M	Representative, Governor	Florida	Official
hillaryclinton	31.5 M	D	F			Personal
ilhan	1.3 M	D	F	Representative	Minnesota	Office
ilhanmn	3.1 M	D	F	Representative	Minnesota	Personal
jerrybrowngov	1 M	D	M	Governor	California	Office
jim_jordan	2.8 M	R	M	Representative	Ohio	Official
joebiden	34.4 M	D	M	VP, President	Delaware	Personal
johnkerry	3.4 M	D	M	Cabinet		Personal
kamalaharris	19.7 M	D	F	Senator, VP	California	Official
laurenboebert	1.4 M	R	F	Representative	California	Personal
leadermcconnell	2.1 M	R	M	Senator	Kentucky	Official
lindseygrahamsc	2.1 M	R	M	Senator	South Carolina	Official
marcorubio	4.4 M	R	M	Senator	Florida	Personal
markmeadows	1 M	R	M	Representative	North Carolina	Official
mattgaetz	1.5 M	R	M	Representative	Florida	Personal
mike_pence	5.8 M	R	M	VP, Governor	Indiana	Personal
mikepompeo	1.4 M	R	M	Representative	Kansas	Personal
mittromney	2.1 M	R	M	Senator	Utah	Personal
nycmayor	1.6 M	D	M	Mayor	New York	Official
nygovcuomo	2.4 M	D	M	Governor	New York	Office
ossoff	1.3 M	D	M	Senator	Georgia	Personal
petebuttigieg	3.6 M	D	M	Cabinet		Personal
potus	22.7 M	D	M	President		Office
presssec	2.4 M	D	F			Office
presssec45	5.8 M	R	F			Office
randpaul	3.9 M	R	M	Senator	Kentucky	Personal
rashidatlaib	1.4 M	D	F	Representative	Michigan	Personal
realbencarson	2.2 M	R	M	Cabinet		Personal
realdonaldtrump	87.4 M	R	M	President		Personal
repadamschiff	3.1 M	D	M	Representative	California	Office
repjohnlewis	1.1 M	D	M	Representative	Georgia	Official
repkatieporter	1.2 M	D	F	Representative	California	Office
repmattgaetz	1.6 M	R	M	Representative	Florida	Official
repmaxinewaters	1.6 M	D	F	Representative	California	Office

repswalwell	1.3 M	D	M	Representative	California	Office
reverend-warnock	1.1 M	D	M	Senator	Georgia	Personal
rondesantis	1.8 M	R	M	Representative, Governor	Florida	Personal
secblinken	1.5 M	D	M	Cabinet		Office
secpompeo	2.9 M	R	M	Representative, Cabinet	Kansas	Office
senfeinstein	1.4 M	D	F	Senator	California	Official
sengillibrand	1.6 M	D	F	Senator	New York	Official
senjohnmccain	2.8 M	R	M	Senator	Arizona	Official
sensanders	12.5 M	I	M	Senator	Vermont	Office
senschumer	3.4 M	D	M	Senator	New York	Office
sentedcruz	2.7 M	R	M	Senator	Texas	Official
senwarren	7.1 M	D	F	Senator	Massachusetts	Office
speakermccarthy	2 M	R	M	Representative	California	Official
speakerpelosi	7.4 M	D	F	Representative	California	Office
speakerryan	3.4 M	R	M	Speaker	Wisconsin	Office
tedcruz	5.2 M	R	M	Senator	Texas	Personal
tedlieu	1.6 M	D	M	Representative	California	Personal
tgowdysc	1.2 M	R	M	Representative	South Carolina	Personal
tulsigabbard	1.5 M		F	Representative	Hawaii	Personal
vp	12.9 M	R	M	Senator, VP	California	Official
vp45	9.9 M	R	M	Vice President		Office
whitehouse	7.1 M	D				Office

This table contains the $N = 70$ political elite Twitter accounts which have over one million followers (as of July 11, 2022). In Party column, D = Democrat, R = Republican, I = Independent. In Gender column, F = Female, M = Male. For accounts that are used by sitting Party, the relevant Party at the time of intervention is used eg. whitehouse is Republican during the time period 2017-2021.

Table C.4: Non-Political Elite Twitter Accounts

Twitter Username	Followers
elonmusk	159.8 M
justinbieber	111.7 M
cristiano	109.9 M
rihanna	108.5 M
katyperry	107.2 M
taylorswift13	94.6 M
arianagrande	85.3 M
ladygaga	83.9 M
kimkardashian	75.3 M
ellendegeneres	75.1 M
selenagomez	66.7 M
billgates	64 M
neymarjr	62.9 M
jtimberlake	61.4 M
imvkohli	58.4 M
britneyspears	5.1 M
shakira	53.8 M
ddlovato	53 M
kingjames	52.7 M
jimmyfallon	50.2 M
bts_twt	48.6 M
mileycyrus	46.5 M
akshaykumar	46.2 M
beingsalmankhan	45.4 M
jlo	44.9 M
iamsrk	43.8 M
bts_bighit	43.6 M
brunomars	42.6 M
oprah	42.1 M
niallofficial	40.4 M
kyliejenner	40.2 M

Twitter Username	Followers
drake	39.4 M
sachin_rt	39.2 M
harry_styles	37.7 M
kevinhart4real	37 M
wizkhalifa	36.3 M
louis_tomlinson	35.3 M
liltunechi	34.4 M
liampayne	33.5 M
iamcardib	32.3 M
ihrithik	32.3 M
kendalljenner	31.9 M
kanyewest	31.7 M
chrisbrown	31.6 M
pink	30.9 M
zaynmalik	30.5 M
khloekardashian	30.3 M
aliciakeys	29.5 M
kaka	29.3 M
nickiminaj	28.1 M
conanobrien	27.7 M
priyankachopra	27.7 M
emmawatson	27.6 M
adele	27.4 M
whindersson	27.1 M
deepikapadukone	26.9 M
aamir_khan	26.8 M
kourtneykardash	26.5 M
m10	26.4 M
shawnmendes	26.1 M
andresiniesta8	25.4 M

We retrieve the list of the top 100 accounts from the social media analytics website Social Blade: `https://socialblade.com/twitter/top/100` retrieved 14 October 2023. We then eliminate the accounts of politicians (including non-U.S. and non-active/past politicians) and organizations. This leaves us with a list of 61 accounts.

Table C.5: News Media Organizations Twitter Accounts

Twitter Username	Followers	Twitter Username	Followers
cnnbrk	63.9 M	SkyNewsBreak	4.9 M
CNN	61.8 M	politico	4.6 M
nytimes	55.1 M	CNNPolitics	4.5 M
BBCBreaking	51.7 M	thehill	4.4 M
espn	49.1 M	BusinessInsider	4 M
SportsCenter	41.9 M	latimes	3.9 M
BBCWorld	40.3 M	guardiannews	3.8 M
TheEconomist	27.2 M	Independent	3.6 M
Reuters	25.7 M	CBCNews	3.5 M
FoxNews	24.3 M	Newsweek	3.5 M
WSJ	20.6 M	nypost	3 M
washingtonpost	20 M	nprpolitics	2.9 M
Forbes	19.7 M	Variety	2.9 M
TIME	19.3 M	MailOnline	2.8 M
ABC	17.8 M	Xnews	2.7 M
AP	16 M	itvnews	2.6 M
BBCNews	15.2 M	Channel4News	2.4 M
cnni	14.7 M	abcnews	2.3 M
SkySportsNews	12.5 M	AJENews	2.2 M
XHNews	11.9 M	PBS	2.1 M
enews	11.7 M	usweekly	2.1 M
TheOnion	11.6 M	BreitbartNews	2 M
guardian	10.9 M	OANN	2 M
HuffPost	10.9 M	TheSun	2 M
TimesNow	10.2 M	CTVNews	1.9 M
BreakingNews	9.4 M	VICE	1.9 M
mashable	9.4 M	foxnewspolitics	1.8 M
NBCNews	9.4 M	YahooFinance	1.7 M
ABSCBNNews	9.1 M	ABCWorldNews	1.6 M
CBSNews	8.9 M	BuzzFeedNews	1.3 M
NewYorker	8.8 M	HuffPostPol	1.3 M
AJEnglish	8.7 M	thedailybeast	1.3 M
NPR	8.7 M	Refinery29	1.2 M
SkyNews	8.4 M	ABCPolitics	1.1 M
people	7.6 M	chicagotribune	1.1 M
TMZ	7.4 M	NewsHour	1.1 M
ARYNEWSOFFICIAL	5.6 M	BBCPolitics	1 M
USATODAY	5 M	YahooNews	1 M

To create this list of all news media organization Twitter accounts with more than one million followers we aggregated several sources including the list of top 100 accounts from Social Blade used for the list of non-political elites, `https://memeburn.com/2010/09/the-100-most-influential-news-media-twitter-accounts/`, `https://viralpitch.co/topinfluencers/twitter/top-200-twitter-influencers/`, and `https://en.wikipedia.org/wiki/List_of_most-followed_Twitter_accounts`. We retrieved the number of followers for each account on 6 December 2023. The final list of 76 accounts include all news media organizations that cover U.S. news stories, in English.

C.0.4 List of Political Identity Words from Rogers and Jones (2021)

activist, all lives matter, all_lives_matter, alllivesmatter, alt-right, anarchist, black lives matter, black_lives_matter, blacklivesmatter, blm, blue lives matter, blue_lives_matter, bluelivesmatter, communist, conservative, democrat, deplorable, feminist, gop, leftist, lgbtq, liberal, libertarian, maga, marxist, men's rights, mens rights, mens_rights, mensrights, nasty woman, nasty_woman, nastywoman, progressive, red pill, republican, socialist, the 99%, the_99%, the99%, woke