Supplemental Information for Weaving it in: How partisan media react to events

Anonymized version

March 12, 2021

Contents

1 Data set

Recording of radio stations started in May 2018, with 72 stations. The number of stations followed increased from there, reaching 175 in August 2018. The number dropped to 155 in March 2019 and remained there until the end of the period analyzed (October 2019). Figure 1 illustrates the number of stations in the data set over time. Of the initial stations, 50 were chosen randomly from the population of US talk radio stations, and more stations were added and dropped from the data set in the course of the next few months according to the interests of the team at [redacted for anonymity]. The total number of US terrestrial talk stations (with registered formats Business News, News/Talk, Public Radio, News, College, Talk) in the covered period was 1897, according to the website Radio-Locator.com.

Table 1 describes the distributions of these stations in terms of content and station subtype, compared to the population of all talk radio stations. Underrepresented station types include stations from the Midwest, college stations, and public radio stations. Underrepresentation of public stations is not likely to be consequential for the final sample of shows. This is because many public stations broadcast the same set of shows, produced for nation-wide broadcasting by public radio networks such as NPR. Figure 2 shows where in the United States the transcribed stations are located.

	Sample %	Population %	t-test p-value
region: Midwest	16	28	0.00
region: Northeast	22	16	0.04
region: South	33	30	0.35
region: West	29	27	0.46
format: Business News	0	1	0.09
format: College	2	13	0.00
format: News	2	2	0.91
format: News/Talk	49	27	0.00
format: Public Radio	32	45	0.00
format: Talk	15	12	0.18

Table 1: Balance table comparing region and format for radio stations in our sample, to the population of US talk radio stations.

The number of distinct radio shows broadcast by these stations added up to 675 by June 2018, and over 1000 by August. Figure 3 shows the total number of shows captured in each week of the observed period.

The speech transcription algorithm was gradually improved in the course of data collection, with the error rate going from 27% in April 2018 to 13% in November 2018, and staying stable since. This error rate was measured using existing transcripts from NPR and Rush Limbaugh, whose audio is likely a little easier to transcribe than the average show. [Reference redacted for anonymization] describe the features of (a subset of) this data set in more detail.

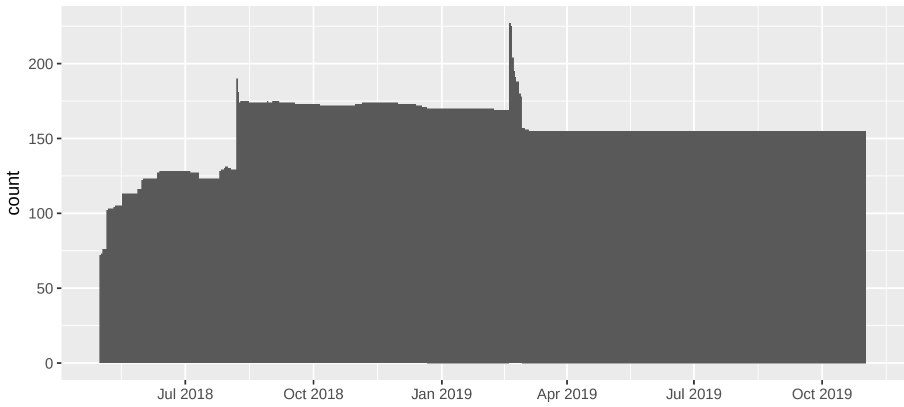

Figure 1: Number of talk radio stations in data set over time.

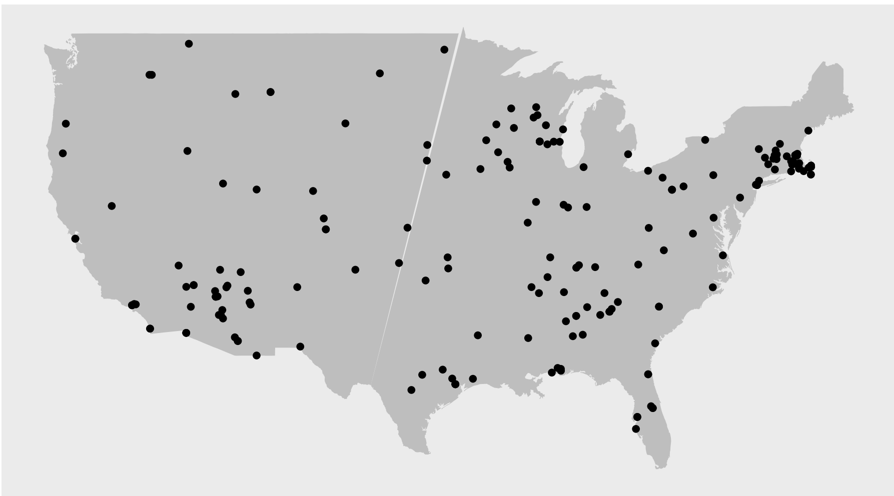

Figure 2: Locations of the continental radio stations in the data set. Two more stations are in Alaska, and one is on Hawaii.

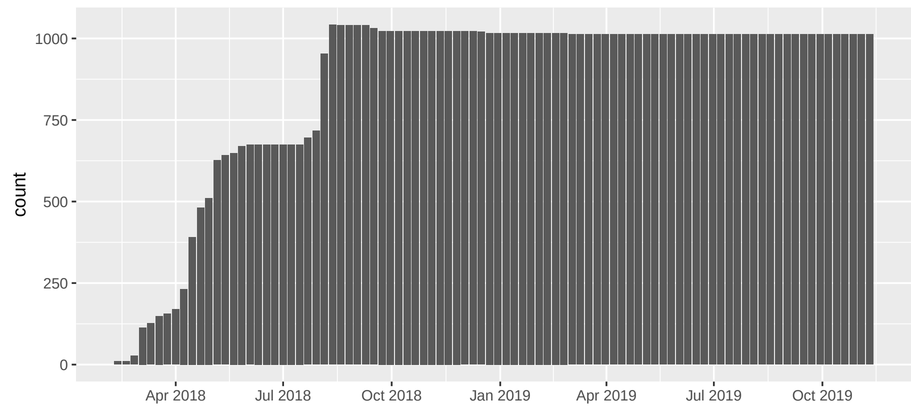

Figure 3:

2 Ascribing mentions to shows

In principle, ascribing a topic mention to a show should be easy. A topic term counts as being part of a show if it was mentioned on a radio station, during a time slot when we know that show is being broadcast on that station. There are two issues with that approach, however. First, the speech-to-text algorithm creating the transcriptions is not perfect. Some topic term mentions are missed, whereas others are false positives. Second, I pooled together radio schedule data (which show is on when) from a range of sources, some more reliable than others. An hour of audio coming from a particular station could have the wrong show label if there is no correct, up-to-date schedule for the station.

Both problems can be remedied when there is more than one airing of the show—that is, when the show is broadcast on more than one of the stations that were recorded that day. I followed the procedure below to decide which topic mentions should be ascribed to each show.

1. For each topic mention, create a "slice" of content that includes up to ten words coming before and after the topic term. For example: "years from now, fifty years, there isn't any evidence of climate change" (aired on KBTK, April 25th 2018).

2. In the entire set of transcripts coming from a particular day, search for clusters of similar mentions (low string distance).

3. For each cluster, do the following:

   (a) Take all mentions in the cluster, and check which show labels they have, based on station scheduling data. For example, the first mention above comes from audio labeled as The Glenn Beck Program, but it is in a cluster together with five (very similar) mentions that are all labeled as being part of the Rush Limbaugh Show.

   (b) For each of the shows that occur at least once as show labels in the cluster, calculate confidence that it is the correct show label for this cluster of mentions. Calculations take into account how many mentions in the cluster have this show label, and also how many times each show was aired on different stations that day, *without* containing a similar mention.

(c) Compare the show label likelihoods, and choose the one with the highest confidence. For example, the mention cluster above was ascribed to The Rush Limbaugh Show with a confidence of .6. This relatively low confidence mostly comes from the fact that there are many airings, labeled in the scheduling data as broadcasts of The Rush Limbaugh Show on April 25th 2018, that did not include any mentions with a content slice similar to this one.

4. Treat each cluster, with its most likely show label, as a single unique mention that happened on that show—but only if its show label confidence is greater than .5. Otherwise, discard the cluster.

3 Mechanical Turk task

Each time one of the political topics was mentioned, the position taken was coded by workers on Amazon's crowdsourcing platform, Mechanical Turk. Workers were allowed to code as many mentions as they wanted, for a payment of $0.14 per mention.

Human coders listened short audio fragments surrounding each mention. This meant that transcription errors were not an issue, and also that the ratings are based on vocal as well are verbal cues. We know from previous work that tone of voice confers unique information (Dietrich *et al.*, n.d., 2019). Audio fragments started 10 seconds before the topic-related phrase (e.g. "global warming") was said, and ended 20 seconds after. After extensive pre-testing, I found that longer fragments very rarely provided information that would change one's initial judgment. Next, the coder was asked to choose between two positions (e.g., "skeptical" or "convinced" about climate change), or "neither" position.

In the case of climate change, for 71% of mentions, the first two coders agreed on the classification. In another 25% of the cases, a third coder broke the tie, and I used the majority opinion as the code for that mention. In the final 4%, all three coders disagreed, and I labeled the fragment "neither". In the case of gun policy, the distribution was: 61% two-coder agreement; 32% two-out-of three majority; 6% no agreement. Immigration fragments were the most difficult to code: the percentages were 54%, 37%, and 9%.

For all three topics, it was rare for two coders to assign opposite positions to the same fragment. In the case of climate change, only 7% of mentions were assigned one position (skeptical or concerned), even though one of the coders suggested the opposite position. In the cases of gun control and immigration, this was 9% and 8%. In other words, coding disagreements are unlikely to have resulted in mentions being assigned the wrong slant. They would, however, have caused some neutral topic mentions to be incorrectly coded as taking a position, and vice versa.

Instructions received minor changes during the coding process, in order to account for common mistakes. Right next to the audio player, coders always saw the following brief instructions on their screen:

Climate:

- Skeptical: climate change evidence is false or unclear, climate change is not an important problem, it is too costly to fight against climate change.
- Concerned: climate change evidence is solid, climate change caused by humans, it is a threat and we need action.
- Neutral: no clear opinion about climate change, and no mention of evidence for or against climate change.

Gun policy:

- Pro-gun: right to own guns, looser gun control laws, guns protect people, second amendment
- Anti-gun: stricter gun control laws, guns cause violence/mass shootings
- Neither: no opinion about gun rights/ gun control, no hints whether the speaker is pro- or anti-gun.

Immigration:

- Supporting immigration: we don't need a wall or more deportations, families should stay together, immigration is good for our country
- Tough on immigration: border needs protection, illegal immigration should be stopped, immigration is bad for our country
- Neither: just news, no opinion about immigration, no hints whether the speaker is supportive or tough.

Finally, coders were encouraged to click through to the longer instructions ("code book") if they were doing the task for the first time, or had not done the task in the past day. The sections below contain the descriptions of each topic position in the final code books.

Climate, Skeptical - The evidence for climate change is false or not certain; predictions did not come true. Scientists are hiding evidence against global warming. The climate is always changing. Humans did not cause global warming. Problems we see today (e.g. wildfires) are not caused by climate change. Even if global warming exists, the effects are not so bad, or they are positive. Climate change is not important compared to other problems. It is too expensive or risky to take action, it would cost too many jobs, it is too soon to take action, it is not our responsibility.

Climate, Concerned - The evidence for climate change is clear. Humans and their greenhouse gas (CO_2) emissions cause global warming. Climate change will have negative effects (e.g. sea levels rising, plants or animals dying) now or later. Problems we see today (e.g. storms, droughts) are due to global warming. We need to act on it (e.g. by using less energy or clean energy). People who are looking for solutions or who are passing climate laws are doing the right thing.

Gun policy, Pro-gun - People have the right to own guns, protected by the second amendment. The government should not take our guns away. There should be fewer laws and rules about owning or buying guns and ammo (e.g. bullets). Gun control does not prevent crimes. People who own guns prevent crimes from happening, because they can defend themselves, their family, and others. People need guns to protect themselves if the government turns against the citizens. Mass shootings are a mental health problem. The US does not have more gun violence because it has more guns.

Gun policy, Anti-gun - There should be stricter rules about who can own and buy guns. People who want to buy a gun should have to pass a background check or get a license. Some types of guns, like assault rifles, should be banned. We need stronger measures to prevent teenagers, or people with mental health problems from having guns. The United States has more gun violence than other countries because it has more guns. Mass shootings would happen less often if it was harder to get a gun.

Immigration, supportive - There should not be a wall on the border with Mexico, and we should deport fewer people. Unauthorized immigrants are often running from violence in their home country. They should be treated well and families should stay together. The rules for legal immigration should not be made stricter. People who were brought into the country as children should be allowed to stay. Immigrants are hard-working, and they contribute to our society. America is a nation of immigrants.

Immigration, tough - We should invest more money and manpower into protecting the border and deporting unauthorized immigrants. If immigrants come or stay here illegally, they broke the law. Immigrants they raise crime rates, they do not pay taxes and should not get government help. We also need stricter policies on legal immigration. Many immigrants don't speak English well, don't adopt American culture, or take jobs from Americans. American-born citizens should come first.

In the code books, coders were encouraged to classify mentions as "neither" if there was not enough context to classify them, if they did not fall into any of the other categories, or if the audio fragment was not about climate change. For example, a piece of news (with no negative or positive tone) about a law that was passed in Congress, or a commercial about climate-proofing your windows. However, the code books also explained that topic mentions can support positions even if the speaker is not giving their own opinion. For example, a news item about new evidence for (or against) climate change would still count as concerned (or skeptical) and not neutral, because its effect could be to make a listener more concerned (or skeptical).

4 Google trends and topic mentions

To define "pre-event" and "post-event" weeks, I use Google Trends data. They give a day-by-day index of the number of Internet searches for a search term

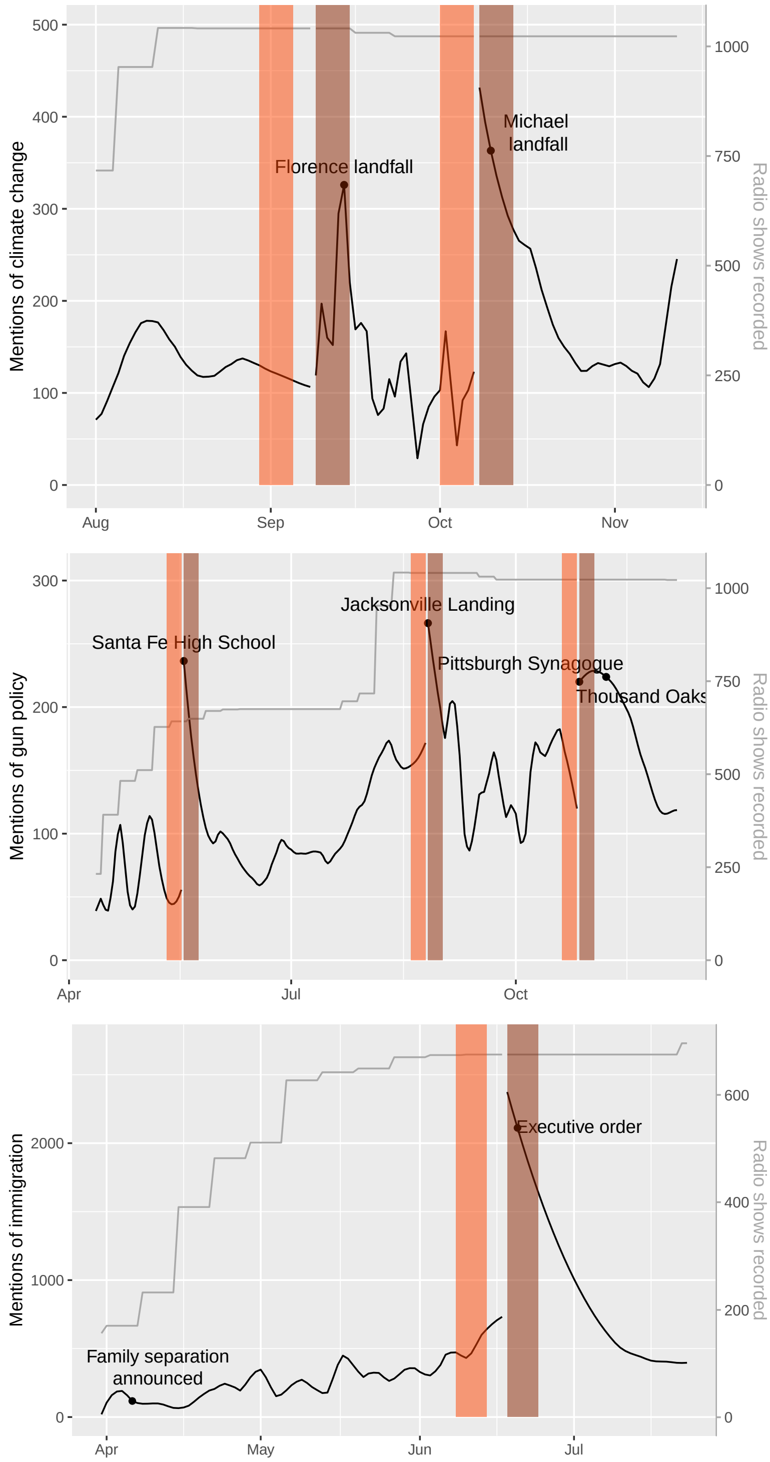

Figure 4: Daily number of talk radio mentions of climate, gun policy and immigration, with significant events. Bands show the pre- (orange) and post- (brown) event weeks as defined by Google Trends. Mention trends smoothed using Loess regression, allowing for discontinuities at the beginning of each post-event week. Light gray lines show the number of talk radio shows recorded on each date.

describing the event (e.g., "hurricane Florence" or "hurricane Michael"). On each day, I compare search activity to the peak-activity day for that event. By my definition, pre-event weeks end on the last day where search activity was less than 5% compared to the peak. Post-event weeks start on the first day where search activity was at least 20% compared to the peak.

Figure 4 shows these periods, along with the number of topic mentions per day on all talk radio shows.[1] While these decision rules do not always line up perfectly with the "before" and "after" of talk radio attention, they do a reasonable job of capturing the baseline and the peak. In the figures, we can also see that the "pre-event" weeks are acceptable baselines, even though no week is ever completely free of (at least local) events that are relevant to these political topics.

5 Classifying radio shows

Table 2 contains all non-political shows that I used to train the political/non-political classifier. Table 3 lists all political shows. For each show, the table includes at least two sources backing its ideological label. A source is considered to confirm an ideological label if it names the program, its host, or another closely associated entity as either "conservative" or "right-wing"; or as "liberal", "left-wing" or "progressive".

Requiring more than one source hedges against the possible ideological bias in ideological bias judgments themselves. For example, the Center for American Progress is itself classified as left-wing by the Media Bias/Fact Check group. I had to leave a number of potentially ideological shows out of this training set because I could find only one source to confirm its slant. I never came across two sources that contradicted each other in their judgments of any given show's bias; when two sources existed, they always agreed.

Two NPR programs (All Things Considered and Morning Edition) were left out of the ideology model's training set for the main analyses, despite the fact that several sources labeled NPR as liberal. Section 8.1 in the paper provides a robustness check that includes these NPR shows in the training set.

In total, the labeled shows had almost 8500 episodes, of which almost 5800 were political. Before training the model on these shows, I held out 10% of each show's episodes, to be used for model testing. To transform the show transcripts into data, I counted and normalized the number of occurrences of 5000 tokenized word pairs in each transcript. In other words, the features fed to the model are term frequency–inverse document frequency (TF-IDF) vectors for 5000 tokenized bigrams. I left out any features whose TF-IDF score was correlated too strongly with any particular show label–for example, hosts' verbal tics, their names or shows sponsors.

Existing literature on categorical ideology classification at the phrase (Iyyer *et al.*, 2014) or document (Yan *et al.*, 2017) level suggests that regularized logistic regression (LR) works well with this amount of data. I tried both LR (with L2 regularization) and Support-Vector Machines (SVMs). I decided between these two models, and tuned both the show-specific feature correlation threshold and the shrinkage parameter c, via blocked k-fold cross-validation. That is, I left out

[1] These are unique mentions, not double-counting mentions on radio shows that are broadcast on more than one station. See Appendix 2 for more on this.

Show	Topic
Food Friday Vox Pop	food
WMT Cooking Show	food
Better Lawns and Gardens	gardening
Classic Gardens and Landscape	gardening
GardenLine w/ Randy Lemmon	gardening
Dr. Bob Martin	health
Purity Products	health
Your Health with Dr. Joe Galati	health
At Home with Gary Sullivan	home
House Talk with Ray Trimble	home
Sturdy Home Improvement	home
Texas Home Improvement	home
Handel on the Law	legal
The Legal Exchange	legal
Your Legal Rights	legal
Financial Advisors with Aubrey Morrow	money advice
Money Matters with Ken Moraif	money advice
The Dave Ramsey Show	money advice
The Financial Exchange	money advice
Afropop Worldwide	music
Afternoon Jazz	music
Classic Jazz with Michele Robins	music
Classical 24 with Andrea Blain	music
Classical 24 with Bob Christiansen	music
Homegrown Music	music
Jesus Christ Show (PRN)	religious
Lutheran Hour	religious
St. John's Lutheran Church	religious
Ben Maller	sports
Buckey Sportsman with Dan Armitage	sports
FOX Sports Radio	sports
Fox Sports Weekends	sports
The Big Sports Show	sports

Table 2: Non-political shows in training set, with their hand-coded topic.

all episodes from the same show at once, and then tried to predict their label with a model trained on the other shows. LR slightly outperformed SVM for the political/non-political model, and SVM slightly outperformed LR for the ideology model.

Once tuned, I tested the final models' performance by using them to label the hold-out episodes of each show, which were all completely new to the model. For each show, when trying to classify its hold-out episodes, I trained a model on all *other* shows. This way, I avoided rewarding the model for making predictions based on show-specific features. The political/non-political LR correctly classified all 50 shows based on their hold-out episodes. The conservative/liberal SVM successfully classified 14 out of 15 political shows. After tuning and testing, I trained the classifiers on the full labeled data sets (training and hold-out).

Conservative

Show	Sources
Ben Shapiro	Media bias/fact check (as The Daily Wire), Politifact, Wikipedia
Glenn Beck	CAP, Pew, Wikipedia
Hugh Hewitt	CAP, Media bias/fact check (as Salem Radio Network News), Wikipedia
Joe Pags	CAP, Wikipedia
Laura Ingraham	CAP, Politifact, Wikipedia
Mark Levin	CAP, Media bias/fact check (as Conservative Review)
Mike Gallagher	CAP, Wikipedia
Rush Limbaugh	CAP, Pew, Politifact, Wikipedia
Sean Hannity	CAP, Pew, Politifact, Wikipedia
The Savage Nation	CAP, Politifact, Wikipedia

Liberal

Show	Sources
Democracy Now!	Media bias/fact check, Wikipedia
Mike Malloy	CAP, Wikipedia, Liberal Talk Radio Wiki
Ring of Fire Radio	Media bias/fact check, Wikipedia
Stephanie Miller	Media bias/fact check (as Fstv), Politifact, Wikipedia, Liberal Talk Radio Wiki
Thom Hartmann	CAP, Media bias/fact check (as Fstv), Politifact, Wikipedia, Liberal Talk Radio Wiki

NPR

Show	Sources
All Things Considered	Media bias/fact check (as NPR), Pew (as NPR)
Morning Edition	Media bias/fact check (as NPR), Pew (as NPR)

Table 3: Political shows in training set, with their ideology label and sources. Sources: CAP (Center for American Progress and Free Press, The Structural Imbalance of Talk Radio, 2007, ampr.gs/2UegLbP); Liberal Talk Radio Wiki (ltradio.fandom.com/wiki/List_of_personalities); Media bias/fact check (mediabiasfactcheck.com); Pew Research Center (journalism.org/interactives/media-polarization); Politifact (politifact.com/personalities/); Wikipedia (wikipedia.com). The entity labeled by the source is in parentheses, if it is something other than the show or its host.

Finally, I applied the political/non-political classifier to all 1005 recorded shows, and the ideology classifier to the 429 shows that were labeled as-political.

There are a few reasons to treat show ideology as binary, rather than continuous. First, existing evidence suggests that radio shows are ideologically sorted, suggesting that it is reasonable to divide shows into a liberal and a conservative group. Second, having two ideology categories is a common choice in studies of talk radio (cf. Yanovitzky & Cappella 2001, Sobieraj & Berry 2011, Center for American Progress and Free Press 2007, Jamieson & Cappella 2008, p.

86). The classification results support this–most political shows can be classified with fairly high confidence as either liberal or conservative, suggesting it is not as important for a model to cover the ideological “middle ground”. Finally, it is more believable to classify the training data into ideological bins. Sources that designate radio shows as right- or left-leaning usually give categorical labels. In a study on TV news, Martin & Yurukoglu (2017) solved this by training a classifier on Congressional speech, with continuous DW-NOMINATE scores as the ideology outcome variable. I found that a domain-adapted binary classifier trained on speeches in the 114th Congress misclassified 3 out of 17 political training shows. Switching to a Congress-based model could thus lead to a significant drop in prediction quality.

6 Modeling mention proportions and counts

For each week on each radio show, I am interested in two outcomes: the number of mentions of a topic; and the proportion of mentions that advocate different positions.

The number of mentions of each topic across shows-weeks has a very skewed, overdispersed distribution. Given that, a linear model of mention counts would have large uncertainty around its coefficients. Moreover, conclusions would be heavily dominated by a handful of shows that have far more mentions than the others. Instead, I use a negative binomial model. The coefficients in this model tell us about the *proportional* change in the outcome variable associated with a change in the predictors.

The full model, which tells us about the differential effect of events on shows with different ideologies, is:

$$\mathbb{E}[Y_{sw}^{count}|T_w, I_s] = exp(\beta_0 + \beta_1 T_w + \beta_2 I_s + \beta_3 T_w I_s + \beta_4 A_s)$$

where Y_{sw}^{count} is the number of topic mentions on show s in week w. It has a negative binomial distribution. T_w indicates whether the week is a pre-event or post-event week, and I_s is the ideological leaning of show s. A_s is the show's airtime in minutes per week. I control for total airtime because shows with more content obviously have more opportunities to mention a topic.

To model position proportions, a so-called fractional response, I use a generalized linear model (GLM) with a logit link function and a quasi-binomial probability mass function (PMF) for the outcome (Papke & Wooldridge, 1996).[2] This is also known as a fractional logit.

The full model is:

$$\mathbb{E}[Y_{sw}^{prop}|T_w, I_s] = logit^{-1}(\beta_0 + \beta_1 T_w + \beta_2 I_s + \beta_3 T_w I_s)$$

where Y_{sw}^{prop} is the fraction of topic mentions on show s in week w that support a particular position (e.g. climate skepticism). This is the proportion among mentions with a position, leaving out mentions coded as “neither”. It has a quasi-binomial distribution.

[2] Using a quasi-binomial PMF instead of a binomial one does not change the estimates but gives us more robust standard errors (Papke, n.d.).

In both models, standard errors are clustered at the show level. Further, many of the quantities of interest are predictions (of mention counts and proportions), and the clustered nature of the sample needs to be reflected in the uncertainty estimates around those quantities. In order to calculate confidence intervals for the predicted quantities (as seen in main text Figures 2 and 3), I use a non-parametric block bootstrap technique (Cameron *et al.*, 2008).

In each iteration of the bootstrap, I sample shows with replacement, extracting the same number of shows as there are in the original data set. For each sampled show, I include all of its observations (show-weeks) in a new simulated data set. I then run the models above on the simulated data and calculate the predicted quantities of interest. Iterating 1000 times results in 1000 predicted quantities, of which the 2.5th and 97.5th quartiles are the bounds of the confidence intervals.

7 Long-term effects: agenda half-lives

Studies of the agenda-setting power of events in mainstream media often find effects that last for months (Lawrence, 2000; Birkland, 2004; Zhang *et al.*, 2017). In this Appendix, I describe how trends in topic mentions tend to evolve, once they have peaked after an event. In other words, I analyze how quickly attention to a topic dissipates. The graphs in section 4 above already give us some visual cues about how long the effects of events can last. Here, I fit a model to the post-event trends.

For each topic, I look at total topic mention counts in the month after each event, day by day. Because I am interested in the downward trend, the start of this month is not the start of the post-event week (which I defined earlier as the first day on which the event reaches some level of social significance). Instead, it is the peak of talk radio attention to the topic: the day with the most mentions.

To avoid catching the beginning of attention to the *next* event, I leave out any days that fall in the next post-event week. This results in six dropped observations for climate change, and six for mass shootings. Finally, for the Pittsburgh Synagogue shooting, I only include the first nine days. This is because ten days after the shooting, there was another mass shooting in Thousand Oaks, California.

After pooling the data across events within topics, I estimate the following simple model:

$$\mathbb{E}[Y_{d,e}^{pct}] = 2^{-\beta d}$$

$Y_{d,e}^{pct}$ is the total number of mentions of the topic on day d after the peak for event e. It is measured as a percentage of peak attention–i.e., attention on day $d = 0$. I did not include an intercept, as $Y_{d=0,e}^{pct} = 1$ by definition. Using base 2 for the exponential decay conveniently allows us to interpret the inverse of β as the half-life of attention; the number of days it takes for mention counts to halve. I estimate the model using non-linear least squares.

This model is not perfect—for instance, attention likely returns to some baseline level in the long run, rather than eventually going to zero. However, it fits the time trend in topic mentions reasonably well.

Figure 5 shows the predicted post-peak attention trend for each topic, alongside the data. The estimated beta coefficients are 0.14 for hurricanes and climate

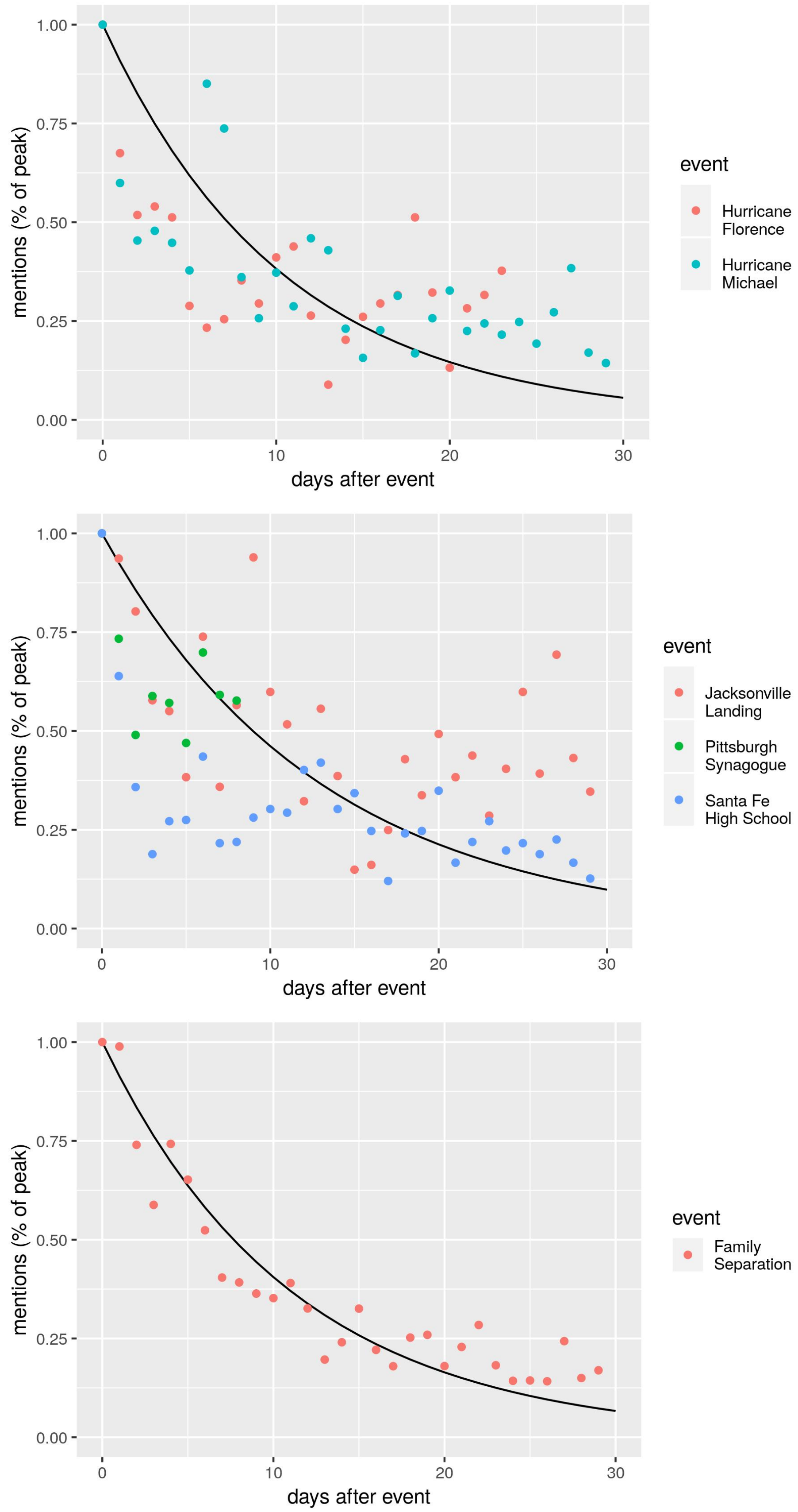

Figure 5: Decline in attention (number of topic mentions) after the peak for each event. Line graphs show the estimated exponential decay of attention. Points show observed attention by event.

change, 0.11 for mass shootings and gun policy, and 0.13 for family separation and immigration. This means that the half-life of attention to these events is 7–9 days.

8 Robustness checks

8.1 News and public radio shows

Most of the existing literature has treated news and public radio shows as separate from political talk radio. However, it turns out that empirically, these shows are not necessarily more "neutral" than the other programs in our data set. In this section, I show that the results are not affected by whether or not news and public radio shows are included.

Among shows classified as political, 25 shows have the word "news" in their name (e.g. "Alabama Morning News"), and 14 shows are produced and distributed by National Public Radio (NPR). First, I look at whether these shows, when they mention a political topic, tend to take neither of the two established positions. Mentions coded as "neither" are usually presentations of facts or straightforward pieces of news about a topic. Bundling all of the observed weeks, on the topic of climate change, the average news show supports neither position in just 16% of its mentions. The same is true for NPR shows. Talking about gun policy, 33% of mentions on the average news show are neutral in this way. On NPR shows it is 34%. Immigration is the topic that invites the most neutral discussion, with 42% of news mentions and 55% of NPR mentions.

A second possibility is that these shows are neutral in the sense that they present both sides of the story equally, for instance by inviting guests with opposite points of view. However, among the topic mentions that have a position, I do not find this type of balance. In the case of climate change, the average news show dedicates more than 89% of its "positioned" mentions to one side of the issue (be it skeptical or convinced). For the NPR shows, it is 95%. On gun policy, these shows spend 73% of their non-neutral mentions arguing for the same side. For the average NPR show, that is 84%. On the topic of immigration, the average news show has 72% of its positioned mentions supporting the same side. For NPR shows, it is 69%. Not surprisingly, all NPR shows tend to pick the same side (in particular, they overwhelmingly are convinced about climate change), whereas the group of news shows is mixed in the direction of their slant. Crucially, none of the numbers above look much different in the sample of non-news, non-NPR shows.

To check the robustness of these findings, I experiment with alternative definitions of news and public radio shows, based on what station(s) broadcast(s) them. All US radio stations have a self-selected format that broadly describes their programming, mostly for marketing purposes. An alternative criterion for news shows would be those shows that are broadcast at least one station with the"All News" format. An alternative criterion for public radio shows would be those shows that are broadcast on at least one station with the "Public Radio" format. These definitions lead to the same conclusion: on news and public radio shows, the discussion of political topics looks no more neutral or balanced than it does on any other political show.

Given these findings, we might view news and public radio shows as simply

		without news, public				with news, public			
		counts		positions		counts		positions	
topic	ideology	pre	post	pre	post	pre	post	pre	post
climate	conservative	0.8	2.1	38	42	0.8	2.2	38	41
climate	liberal	2.2	4.1	92	95	2.0	4.1	93	95
gun policy	conservative	1.9	3.3	21	29	2.0	3.3	22	30
gun policy	liberal	0.7	1.0	71	77	0.7	1.1	68	79
immigration	conservative	4.3	25.3	70	67	5.1	27.4	70	66
immigration	liberal	7.3	14.5	31	27	8.2	17.7	30	28

Table 4: Predicted mention counts and positions (percentage “convinced”, “anti-gun” and “tough on immigration” positions), pre- and post-event, for each political topic, without and with NPR shows or news shows.

another type of talk radio with political content. For that reason, I repeat the analyses, including shows with “news” in the name and shows produced by NPR. I also add two NPR programs to the liberal show training set for the ideology classifier. This improves performance: testing the model on unseen episodes, the classifier now correctly guesses the ideology of *all* shows. Table 4 shows the results, alongside the original ones without NPR and news shows. We can see that the basic thrust is the same.

8.2 Show classification thresholds

In the analyses above, shows are classified based on 50% thresholds: they are labeled political, and conservative, if classifiers assign them a 50%-or-higher probably of being so. However, the training set for each model is a simply set of shows that can reliably be labeled as non-political, liberal or conservative. This set probably does not reflect the actual balance of show ideologies in the full sample. It is likely, then, that the models’ intercept estimates are biased. Moreover, perhaps not all political shows are slanted: it is possible there are moderate shows in the sample, which I am unjustly labeling as ideological.

Figure 6 shows the results of the show classification effort. It looks like the choice of ‘politicalness’ threshold could be important, because some shows are in fact difficult to classify. Only 70% of shows can be labeled as political with at least 70% certainty. In terms of ideology, the picture looks somewhat more robust. 80% of political shows get an ideological label with over 70% certainty. Nonetheless, we may be interested in how results change if we exclude shows whose ideological class is unclear.

Here, I repeat the key analyses, varying my decisions about show classes in two ways. First, I move the political decision threshold above or below 50%, biasing the model towards labeling fewer or more shows as political. Second, I create bands around the ideology threshold, excluding shows that the model is uncertain about. For example, I might only include shows for which the classifier is at least 60% certain that they are either liberal or conservative. Table 5 shows the results of the former analysis. Table 6 shows the latter. Neither decision changes the results in any significant way, except that stricter ‘politicalness’ thresholds lead to somewhat more topic mentions at baseline. This makes sense, since I am excluding shows that spend less time covering political topics.

		climate				gun policy				immigration			
		counts		positions		counts		positions		counts		positions	
threshold	ideology	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post
0.4	conservative	0.7	1.9	38	45	1.6	2.9	22	29	3.8	23.3	66	63
0.4	liberal	1.9	3.7	92	95	0.6	0.8	69	78	6.4	12.5	28	25
0.5	conservative	0.8	2.1	38	42	1.9	3.3	21	29	4.3	25.3	70	67
0.5	liberal	2.2	4.1	92	95	0.7	1.0	71	77	7.3	14.5	31	27
0.6	conservative	0.8	2.3	37	41	2.0	3.5	20	29	4.3	25.8	70	68
0.6	liberal	2.4	4.6	91	95	0.7	1.2	72	78	5.2	16.6	28	30

Table 5: Predicted mention counts and positions (percentage “convinced”, “anti-gun” and “tough on immigration” positions), pre- and post-event, for each political topic. Threshold indicates level of certainty we need in order to call a show “political” and include it in the data set.

		climate				gun policy				immigration			
		counts		positions		counts		positions		counts		positions	
threshold	ideology	pre	post	pre	post	pre	post	pre	post	pre	post	pre	post
0.5	conservative	0.8	2.1	38	42	1.9	3.3	21	29	4.3	25.3	70	67
0.5	liberal	2.2	4.1	92	95	0.7	1.0	71	77	7.3	14.5	31	27
0.6	conservative	0.8	2.1	37	41	1.9	3.4	21	29	4.4	25.8	70	67
0.6	liberal	2.4	4.4	93	96	0.7	1.0	75	75	8.0	14.8	31	27
0.7	conservative	0.8	2.1	35	40	2.0	3.5	20	29	4.2	25.2	71	69
0.7	liberal	2.5	5.9	96	97	0.9	1.2	78	81	5.5	17.0	31	26

Table 6: Predicted mention counts and positions (percentage “convinced”, “anti-gun” and “tough on immigration” positions), before and after events, for each political topic. Threshold indicates level of certainty we need in order to call a show “conservative” or “liberal”, and to include it in the data set.

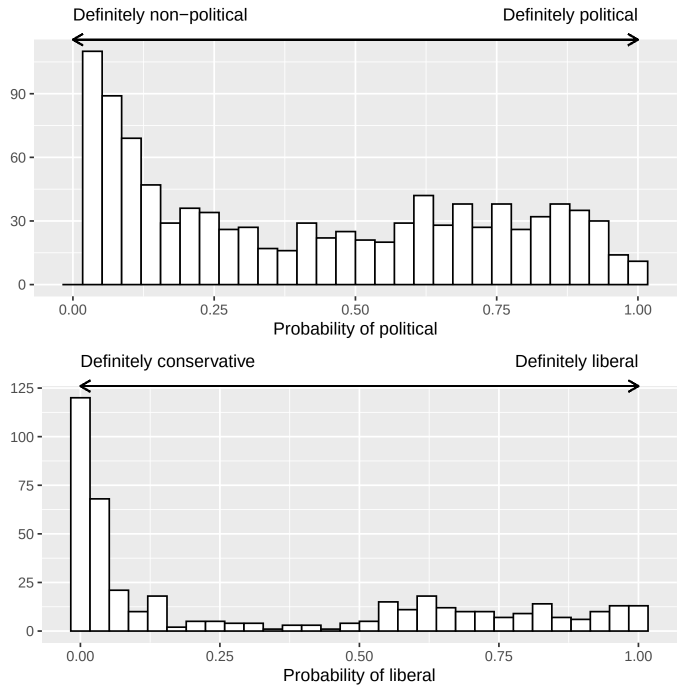

Figure 6: Distribution of prediction probabilities for shows from politicalness and ideology classifiers.

References

Birkland, Thomas A. 2004. "The World Changed Today": Agenda-Setting and Policy Change in the Wake of the September 11 Terrorist Attacks. *Review of Policy Research*, **21**(2), 179–200.

Cameron, A Colin, Gelbach, Jonah B, & Miller, Douglas L. 2008. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, **90**(3), 414–427.

Center for American Progress and Free Press. 2007. *The Structural Imbalance of Political Talk Radio.*

Dietrich, Bryce J, Hayes, Matthew, & O'Brien, Diana Z. Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech. *American Political Science Review*, 1–22.

Dietrich, Bryce J, Enos, Ryan D, & Sen, Maya. 2019. Emotional arousal predicts voting on the US supreme court. *Political Analysis*, **27**(2), 237–243.

Iyyer, Mohit, Enns, Peter, Boyd-Graber, Jordan, & Resnik, Philip. 2014. Political ideology detection using recursive neural networks. *Pages 1113–1122 of: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).*

Jamieson, Kathleen Hall, & Cappella, Joseph N. 2008. *Echo chamber: Rush Limbaugh and the conservative media establishment.* Oxford University Press.

Lawrence, Regina G. 2000. *The politics of force: Media and the construction of police brutality.* Univ of California Press.

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

Papke, Leslie E. *Papke and Wooldridge (1996) flogit Instructions.*

Papke, Leslie E, & Wooldridge, Jeffrey M. 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics*, **11**(6), 619–632.

Sobieraj, Sarah, & Berry, Jeffrey M. 2011. From incivility to outrage: Political discourse in blogs, talk radio, and cable news. *Political Communication*, **28**(1), 19–41.

Yan, Hao, Lavoie, Allen, & Das, Sanmay. 2017. The Perils of Classifying Political Orientation From Text. *Linked Democracy: Artificial Intelligence for Democratic Innovation*, **858**, 8.

Yanovitzky, Itzhak, & Cappella, Joseph N. 2001. Effect of Call-In Political Talk Radio Shows on Their Audiences: Evidence from a Multi-Wave Panel Analysis. *International Journal of Public Opinion Research*, **13**(4), 377–397.

Zhang, Yini, Wang, Yidong, Foley, Jordan, Suk, Jiyoun, & Conathan, Devin. 2017. Tweeting Mass Shootings: The Dynamics of Issue Attention on Social Media. *Page 59 of: Proceedings of the 8th International Conference on Social Media & Society.* ACM.