Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings

Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou,
Matthew Gentzkow, Jesse Shapiro, Dan Jurafsky

By: Laura Lengre

Introduction

Published: April, 2019

Why is this research important:

- Uses different devices and identifies different methods of polarization
- Looks into more cohesive topics
- Sheds light on how group divisions manifest in language (tweets) and computational methods
- Builds a more comprehensive framework for studying linguistic aspects of polarization in social media

Introduction

- Highly polarized politically (Elites, political parties, media in the US)
- Studying 4.4 million tweets on 21 mass shootings (Twitter)
- Deeper understanding of Group Divisions (Democratic and Republican groups) through language
- Data collection: list of mass shootings from 2015-2018 (Gun Violence Archive)

Lexical Method

- Clustering of tweet embeddings
- Human evaluation shows that their approach generates more cohesive topics than traditional LDA-based models

Background: Previous Studies

- Literature studies focused on the role of media and politicians
- Prior NLP work has considered how to extract factual info about gun violence from news
- Quantify stance and public opinion on Twitter and across the web

Now we're advancing NLP approaches to public discourse surrounding gun violence by introducing methods to analyze other linguistic manifestations of polarization

Data collection and Partisan Assignment:

- List of mass shootings from 2015 - 2018 (Gun Violence Archive)
- Built list of relevant tweets for the 2 weeks following the event
- Tweet is relevant only if contains: location-based representative keyword & 1 lemma from following list (“shoot”, “gun”, “kill”, “attack”, “massacre”, “victim”)
- Filtered out retweets and deactivated user’s tweets

Party affiliation of users: estimated from political accounts they followed

Event city / town	State	Specific location	Date	No. victims	Race / ethnicity of shooter	No. tweets	No. partisan tweets	No. Dem tweets	No. Rep tweets
Chattanooga	TN	Military Recruitment Center	7/16/15	7	Middle Eastern	29573	20709	5925	14784
Roseburg	OR	Umpqua Community College	10/1/15	18	Mixed	18076	11505	6419	5086
Colorado Springs	CO	Planned Parenthood clinic	11/27/15	12	White	55843	39719	26614	13105
San Bernardino	CA	Inland Regional Center	12/2/15	35	Middle Eastern	70491	45819	20798	25021
Kalamazoo	MI	multiple	2/20/16	8	White	10986	6807	4350	2457
Orlando	FL	Pulse nightclub	6/12/16	102	Middle Eastern	1831082	872022	450784	421238
Dallas	TX	Black Lives Matter protest	7/7/16	16	Black	260377	144205	64628	79577
Baton Rouge	LA	streets	7/17/16	6	Black	46126	29015	12019	16996
Burlington	WA	Cascade Mall	9/23/16	5	Middle Eastern	8171	4993	1838	3155
Fort Lauderdale	FL	Fort Lauderdale airport	1/6/17	11	Hispanic	12525	7194	3073	4121
Fresno	CA	downtown	4/18/17	3	Black	8868	6128	1377	4751
San Francisco	CA	UPS store	6/14/17	5	Asian	10487	6627	4346	2281
Vegas	NV	Route 91 Harvest Festival	10/1/17	604	White	1286399	726739	315343	411396
Thornton	CO	Walmart	11/1/17	3	White	14341	9170	5527	3643
Sutherland Springs	TX	Texas First Baptist Church	11/5/17	46	White	154076	106220	52513	53707
Parkland	FL	Marjory Stoneman Douglas High School	2/14/18	31	White	272499	186570	113856	72714
Nashville	TN	Waffle House	4/22/18	8	White	38680	24326	14606	9720
Santa Fe	CA	Santa Fe High School	5/18/18	23	White	73621	42968	26784	16184
Annapolis	MD	Capital Gazette	6/28/18	7	White	27715	18468	11863	6605
Pittsburgh	PA	Tree of Life Synagogue	10/27/18	18	White	59925	36920	22735	14185
Thousand Oaks	CA	Borderline Bar and Grill	11/7/18	23	White	117815	62812	40328	22484

Table 3: Data properties.

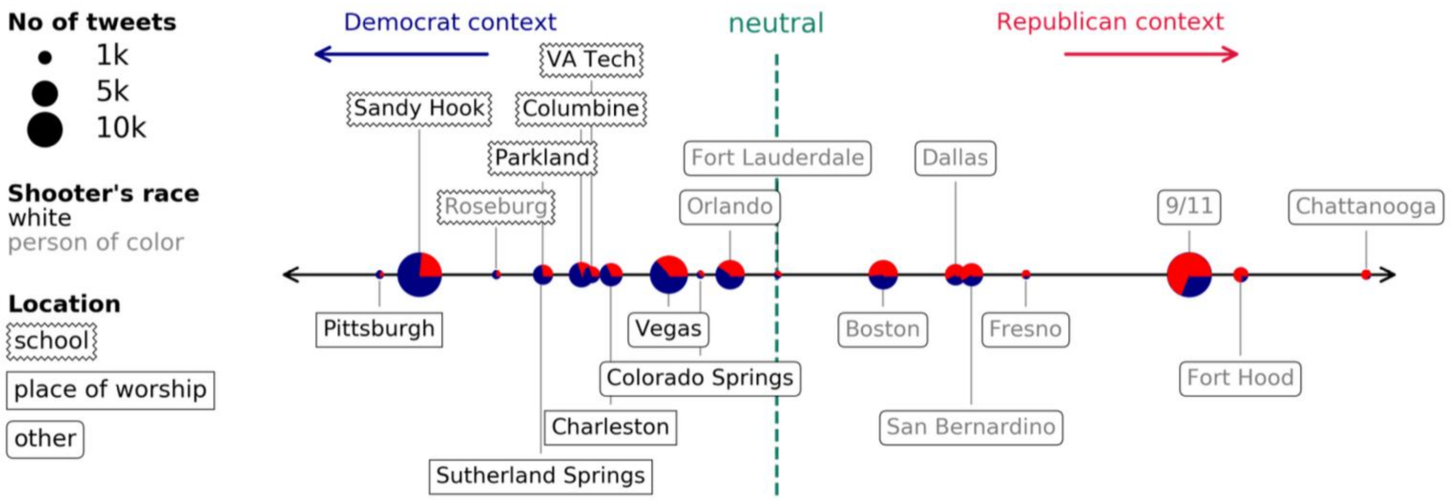


Figure 9: The partisanship of events of mass violence when used as a context for a given mass shooting. The position of the events on the line represents their partisan log odds ratio (Democrat $<$ 0 (neutral) $<$ Republican). The pie charts indicate the proportion of Democrat and Republican users' tweets that involve this "context" event.

NLP framework to uncover 4 linguistic dimensions:

1. **Topic Choice** = tool for agenda-setting, establishing what an author/institution deems worth of discussion (location, setting, demographics of the shooter and victims)
2. **Framing** = grounding and contrasting terms like “terrorist” and “crazy” (Dem and Rep using these terms differently)
3. **Affect** = ideological reasoning (emotional expression)
4. **Illocutionary force** = necessity and possibility (used usually for calls for change/action and express mental state about the event)

Research Questions

1. What language and emotions from tweets correlate with Republicans and Democrats?

2. What phenomenon contributes to polarization?

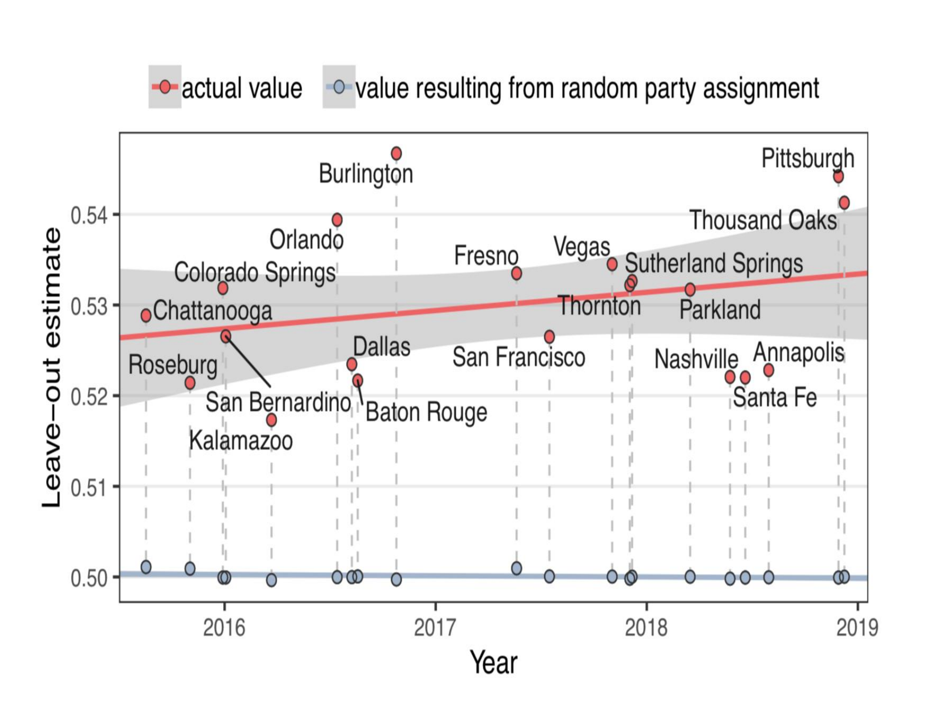
Methodology: Quantifying Overall Polarization

- Pre-processing = built a vocabulary for each event
- Each vocab contains unigrams & bigrams (tokens)
- tokens counted after stemming via NLTK's SnowballStemmer and stop word removal

Snowball Stemmer = Stemming reduces words to their base form (removing suffixes and prefixes), allows text analysis to treat related words uniformly and improves search

Findings - Overall Polarization

- Discussion of each event is highly polarized
- Values range (0.517 to 0.547)
- Measured by leave-out estimate of phrase partisanship (Gentzkow et al., forthcoming)
- Post-event polarization



Methodology: Topics and Framing

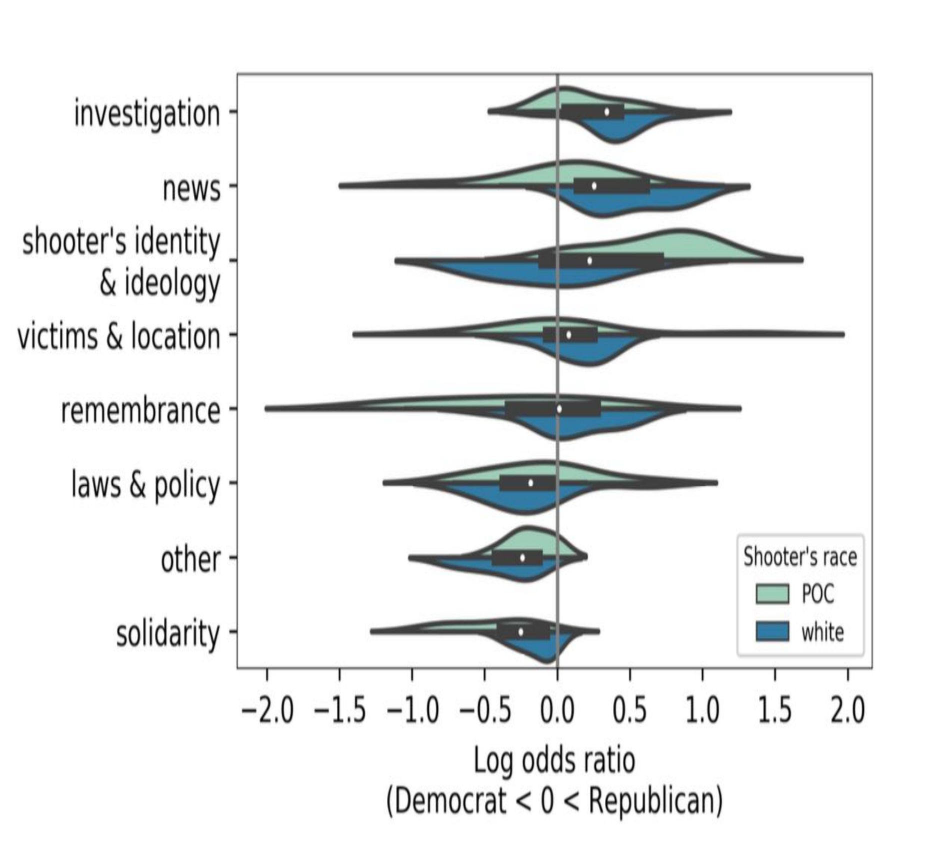
- Topic assignment
- Own model outperforms the LDA- based methods with word intrusion and tweet intrusion

Topic	10 Nearest Stems
news (19%)	break, custodi, #breakingnew, #updat, confirm, fatal, multipl, updat, unconfirm, sever
investigation (9%)	suspect, arrest, alleg, apprehend, custodi, charg, accus, prosecutor, #break, ap
shooter's identity & ideology (11%)	extremist, radic, racist, ideolog, label, rhetor, wing, blm, islamist, christian
victims & location (4%)	bar, thousand, california, calif, among, los, southern, veteran, angel, via
laws & policy (14%)	sensibl, regul, requir, access, abid, #gunreformnow, legisl, argument, allow, #guncontolnow
solidarity (13%)	affect, senseless, ach, heart, heartbroken, sadden, faculti, pray, #prayer, deepest
remembrance (6%)	honor, memori, tuesday, candlelight, flown, vigil, gather, observ, honour, capitol
other (23%)	dude, yeah, eat, huh, gonna, ain, shit, ass, damn, guess

Table 1: Our eight topics (with their average proportions across events) and nearest-neighbor stem embeddings to the cluster centroids. Topic names were manually assigned based on inspecting the tweets.

Findings

- Republicans: *Investigation, News,* *and* shooter's *Identity & Ideology*
- Democrats: *Laws & policy and Solidarity*
- Rep topics relate more to the shooter
- Dem topics relate more to victims



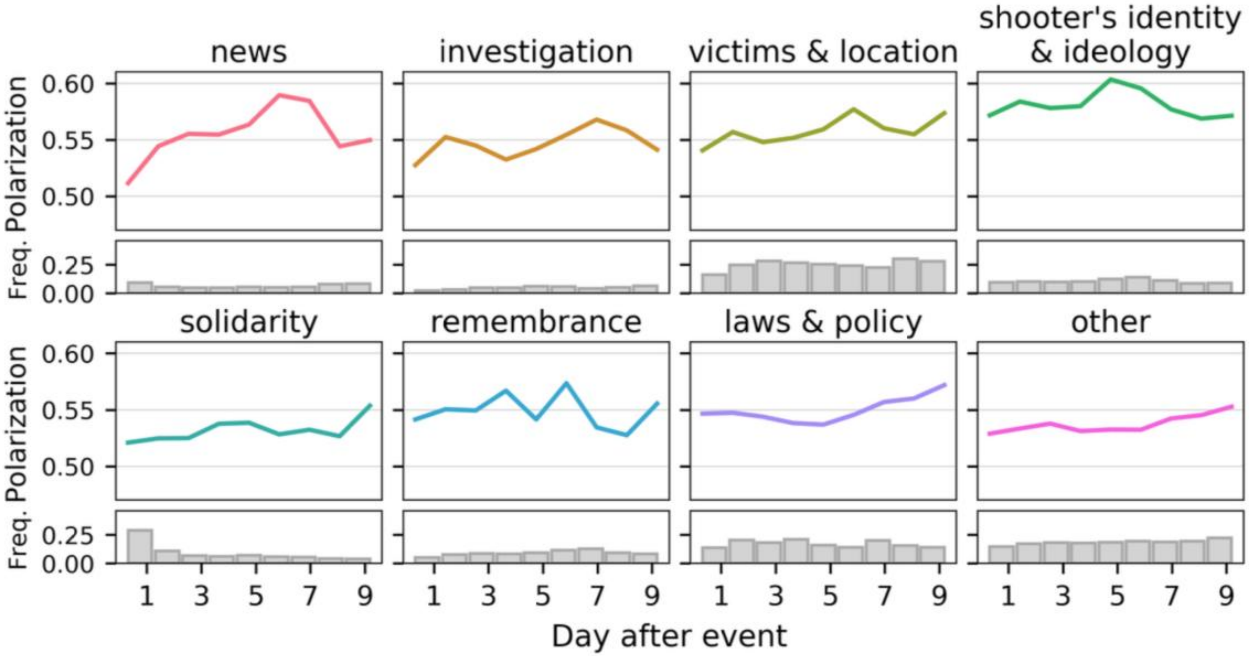


Figure 6: Las Vegas within-topic polarization in the days after the event. The bar charts show the proportion of each topic in the data at a given time.

Methodology: Specific Framing Devices

- Partisan tokens
- Grounding: study whether there is polarization in which prior events to reference certain mass shootings
- Past *context events (Sandy Hook and 9/11)*
- Calculated the partisan log odds ratio

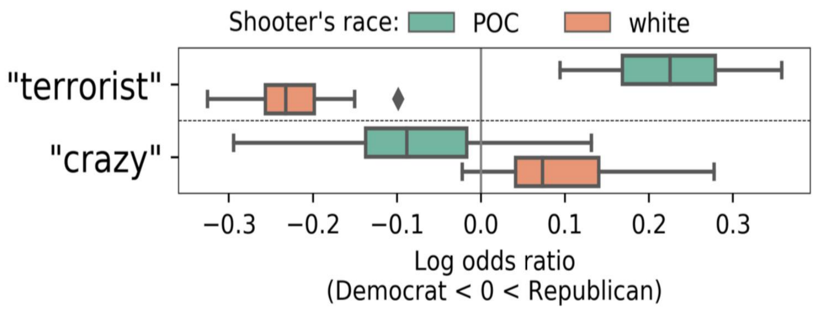


Figure 8: The log odds ratios of "terrorist" and "crazy" across events, grouped by the shooter's race. The boxes show the interquartile range and the diamond an outlier.

Findings

- Focus on partisanship of “terrorist” and “crazy”
- Shooters race
- Differences of using these terms b/w Democrats and Republicans
- Shows how shooter’s race influences how people conceptualize a certain event

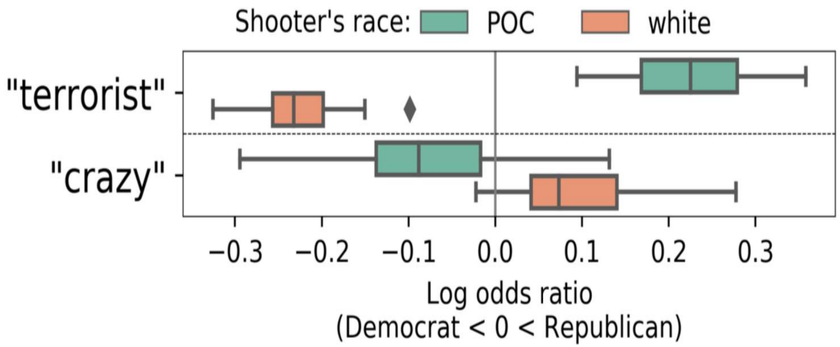


Figure 8: The log odds ratios of “terrorist” and “crazy” across events, grouped by the shooter’s race. The boxes show the interquartile range and the diamond an outlier.

Methodology: Affect

Democrats use lexicons that are associated with positive sentiments, sadness, and trust

Republicans use fear and disgust, especially if the shooter is a person of color (POC)

Anger, trust, and negative sentiments expressed by both parties

Conservatives score higher than liberals based on subjective measures of fear

Disgust sensitivity: associated with political Conservatism

E Emotion Lexicon

The following words were the final stems in our emotional lexicon.

positive love, friend, pray, thought, affect, bless, god, pleas, communiti, hope, stand, thank, help, condol, will, comfort, time, strong, work, support, effect, strength, feel, peac, word, rest, give, great, action, good

negative hate, violenc, hatr, of, evil, tragedi, will, word, attack, sad, feel, anger, murder, shoot, massacr, want, need, pain, kill, griev, crime, ignor, victim, lost, grief, senseless, tragic, fear, loss, sick

sadness senseless, loss, tragedi, lost, devast, sad, love, griev, horrif, terribl, pain, violenc, condol, broken, hurt, feel, victim, mourn, horrifi, will, grief, ach, suffer, sick, kill, aw, sicken, evil, massacr, mad

disgust disgust, sick, shame, ignor, wrong, blame, hell, ridicul, idiot, murder, evil, coward, sicken, feel, disgrac, slaughter, action, bad, insan, attack, pathet, outrag, polit, terrorist, mad, damn, lose, shit, lie, asshol

anger gun, will, murder, kill, violenc, wrong, shoot, bad, death, attack, feel, shot, action, arm, idiot, crazi, crimin, terrorist, mad, hell, crime, blame, fight, ridicul, insan, shit, die, threat, terror, hate

fear danger, threat, fear, arm, gun, still, shooter, attack, feel, fight, hide, murder, shot, shoot, bad, kill, chang, serious, violenc, forc, risk, defend, warn, govern, concern, fail, polic, wrong, case, terrorist

trust school, like, good, real, secur, show, nation, don, protect, call, teacher, help, law, great, save, true, wonder, respons, sad, answer, person, feel, safe, thought, continu, love, guard, church, fact, support

The following words were used as seeds to generate this lexicon, as described in the main text.

positive love, donat, heart, thought, strength, bless, solidar

negative hatr, hate, griev, grief, wrong

sadness mourn, sadden, griev, grief, sad, suffer, affect, broken, senseless, loss, heartbroken

disgust disgust, disgrac, shame, gut, slaughter, sicken, sick, ill, lunat, coward

anger deserv, lynch, gang, threat, mad, sicken, harm, enforc, firearm, ridicul, assault

fear risk, hide, danger, warn, fear

trust secur, coach, safe, hero, nation

Findings

Democrats use lexicons that are associated with positive sentiments, sadness, and trust

Republicans use fear and disgust, especially if the shooter is a person of color (POC)

Anger, trust, and negative sentiments are likely to be expressed by both parties

Conservatives score higher than liberals based on subjective measures of fear

Disgust sensitivity: associated with political Conservatism

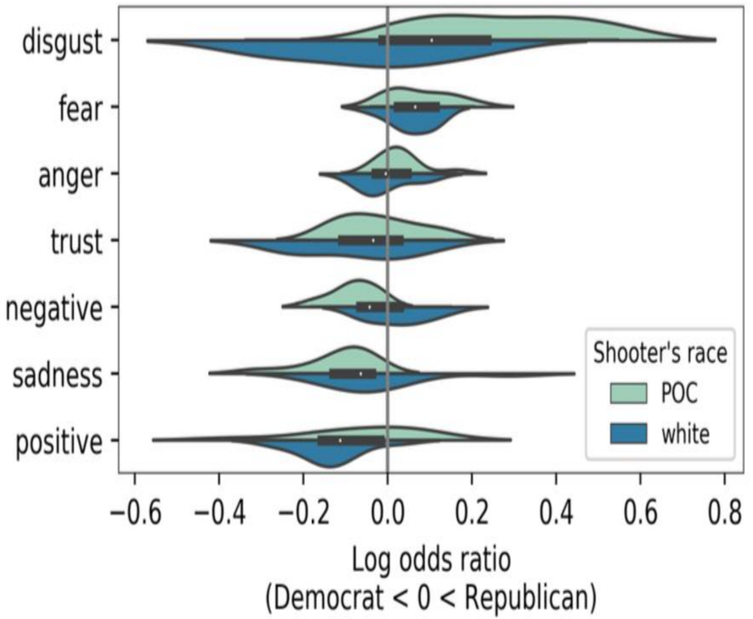


Figure 10: The log odds of each emotion category in our lexicon (one observation represents one event).

Methodology - Modality and Illocutionary Force

- Modals
- Necessity and Possibility
- Should, Must, Have to, Need to

This roller coaster debate **MUST** STOP! Sensible gun ownership is one thing but assault weapons massacre innocent lives. The savagery of gore at #Parkland was beyond belief & **must** be the last.
In times of tragedy **shouldn't** we all come together?! Prayers for those harmed in the #PlannedParenthood shooting.
Communities **need to** step up and address white on white crime like the Las Vegas massacre. White men are out of control.
he BLM protest shooting, planned parenthood, now cali... domestic terrorism will crumble this country, SANE PPL **HAVE TO** FIGHT BACK
Shooting cops is horrible, cannot be condoned. But **must be** understood these incidents are outgrowth of decades of police abuses. #BatonRouge
1. Islamic terrorists are at war with us 2. Gun free zones = kill zones 3. Americans **should be** allowed to defend themselves #Chattanooga
Las Vegas shooting Walmart shooting and now 25 people killed in Texas over 90 people killed Mexico **should** build that wall to keep the US out
CNN reporting 20 dead, 42 injured in Orlando night club shooting. Just awful. The US **must** act to control guns or this carnage will continue.

Table 2: Random sample of tweets with modals. Only one of the eight (Ex. 6) tweets expresses ideas traditionally associated with conservative ideology.

Findings

- 200 modal uses: ~78% e express calls for change/action & ~40% express user's mental state
- Modals are over-represented in the *laws & policy* topic

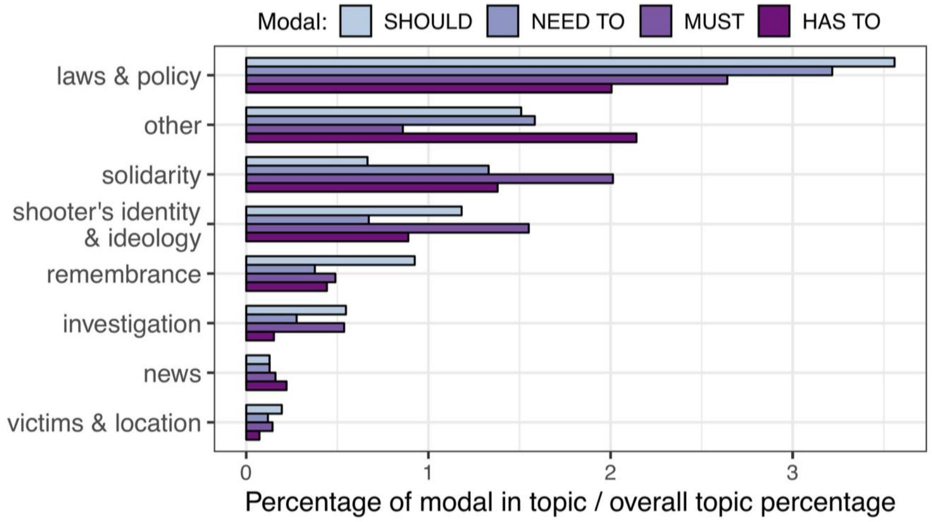


Figure 11: The representation of modals in each topic. Values represent averages across events.

My Commentary

Although race was an important factor brought up, gender and age wasn’t looked into despite having tokens about gender being mentioned in some listed tweets/posts

There were many different categories and topics that were researched that brought different perspectives on polarization on social media and group divisions. This research doesn’t only look into one or two methodologies, but multiple, making it more interesting

Quiz Time!

Quiz Question 1
What are the 4 linguistic dimensions used?

A) Topic choice, framing, affect, and reasoning

B) Topic choice, framing, affect, and illocutionary force

C) Topic choice, formatting, affect, illocutionary force

D) Role, event specific, illocutionary force, affect

Quiz Question 2
Which categories do Republicans usually talk about when the shooter is White?

A) News & Ideology

C) Laws and Policy & Investigations

B) News & Investigation

D) Investigation & Ideology

Quiz Question 3
T/F: Democrats were more focused on the conditions of the shooter while Republicans focused more on the victims.

A) True

B) False

Quiz Question 4
What was the Republicans' most expressed emotion within the collected tweets?

A) Sadness

B) Disgust

C) Anger

D) Trust

Questions?

Thank you!