10-4-2022

Year 2 Final Report: Project Performance Reporting July 1, 2021- June 20, 2022

Matthew L. Jensen

Shaila Miranda

Shane Connelly

Ares Boira Lopez

Cecilia Gordon

See next page for additional authors

Authors

Matthew L. Jensen; Shaila Miranda; Shane Connelly; Ares Boira Lopez; Cecilia Gordon; Marina Mery; Divya Patel; Bachazile Sikhondze; Joseph Stewart; and National Counterterrorism Innovation, Technology, and Education

Year 2 Final Report

Project Performance Reporting July 1, 2021 — June 30, 2022

Primary Investigator:
Matthew L. Jensen

Co-Primary Investigators:
Shaila Miranda
Shane Connelly

Student Investigators (alphabetical):
Ares Boira Lopez, Cecilia Gordon, Marina Mery,
Divya Patel, Bachazile Sikhondze, Joseph Stewart

University of Oklahoma

October 4, 2022

Contents

Executive Summary

This project examines messaging strategies on publicly accessible microblogs (e.g., Twitter) used by extremist ideological groups. Our objective is to provide Department of Homeland Security (DHS) decision-makers and associated partners with insights about processes extreme ideological groups use to recruit members, harness social identities, mobilize communication around issues, increase commitment to extremism, and incite violent action. We analyze digital traces (e.g., websites, microblog archives) and conduct controlled, randomized experiments to understand how messaging content and strategies foreshadow extreme cognitions, affect, and behaviors. Key insights from our analyses have uncovered the following insights:

Key Findings from Digital Trace Results
- Rise in religious rhetoric on microblogs preceded violent events. We observed this phenomenon across multiple jihadist attacks. A similar, though more muted, rise in religious rhetoric preceded the Jan 6 Capitol riots.
- Violent ideological groups use appeals to social identity along with language that justifies the group's stances and emphasizes differences with outgroups. Non-violent groups use appeals to social identity along with language that focuses on group agency, future possibilities, and is more hesitant.

Implications: These findings provide important signals for analysts monitoring rhetoric from known and emerging ideological groups that mark escalation toward extremism and violence. Findings also identify key language differences between non-violent and violent groups.

Key Findings from Experiment Results
- Simple exposure to a social media feed on an ideological topic can change how strongly people feel about the topic and can move them more towards supporting one side of the issue.
- Content that includes social identity influences attitudes, recall of social media tweets, likelihood of mentioning specific tweets in a response, and tweet dissemination in complex ways.
  - Social media feeds that mention racial social identities cause more negative emotions (e.g., aggression) in those who read and respond to these messages than feeds that mention political social identities. This suggests that appeals to inherent social identities relate more strongly to negative emotions than appeals to chosen social identities.
  - Generally, with inherent social identity content, tweets supportive of that identity were recalled, mentioned, and shared with others who may oppose the view to a greater extent. However, when content about other chosen social identities was also included, recall and dissemination were suppressed. This suggests that referring to more than one type of social identity makes the tweets harder to remember and reduces other people's desire to share them.
- Coherence of a message feed (single issue focus versus multiple issue focus) has several important effects such as stronger attitudes that align with the ideology, higher

personal importance of the issue, and increased perceived credibility of the feed and the ideological group that posted it.

- However, high coherence also resulted in less overall recall of specific message content aligned with the group's purpose than low coherence.

Implications: First, exposure to controversial messages will push people further to one or the other side of an issue. As analysts monitor the publics' engagement with controversial messages, level of exposure may inform extremity of stance. Additionally, extremist groups or individuals who post frequently will likely exert more influence than those who post less regularly.

Second, when groups invoke inherent social identities, their messaging will be particularly impactful in stirring up negative affect, increasing recall, and widening dissemination. However, invoking other chosen social identities in messaging appears to disrupt appeals to inherent social identities. This observation may be an important step toward mitigation of extreme online rhetoric: When a group invokes one social identity, reminding or invoking other social identities may blunt the messaging effects.

Finally, coherence, as a messaging strategy, has many positive effects for the groups using it. Analysts should pay particular attention to groups employing focused messaging around core issues. Such messaging is likely to be effective in influencing followers and observers.

Introduction

This project examines messaging strategies on publicly accessible microblogs (e.g., Twitter) used by extremist ideological groups. These groups harness the "weak ties" arising from the asymmetry in how ties form on microblogs toward rapid and widespread message diffusion. Though such groups may use other covert media for communication, publicly accessible microblogs such as Twitter permit them to coopt neutral observers into groups that the public may not immediately recognize as extremist.

Our objective is to provide Department of Homeland Security (DHS) decision-makers and associated partners with insights about processes extreme ideological groups use to recruit members, harness social identities, mobilize communication around issues, increase commitment to extremism, and incite violent action. Iterating between analyzing extremist microblog archives and lab experiments, our research team is systematically examining messaging content and strategies that foreshadow extreme cognitions, affect, and behaviors. We focus on English language content most germane to understanding domestic terrorism incidents that may occur within the U.S. Our work provides insights into how messaging content and strategies promoting and foreshadowing violence can be detected, and threats thereby disrupted.

Summary of Year 2 Effort

During year 2, we engaged in two streams of data collection and analysis. The first was the study of trace data and the second was a series of experiments. With the study of trace data, we focused on digital traces left by ideological groups and we attempted to answer the following questions:

- How do violent and non-violent ideological groups use cognitive and affective processes in messaging?
- What dissemination tactics do violent ideological groups use prior to and following violent incidents?
- What justifications do violent and non-violent groups use for their actions?
- What are the types and timings of crimes committed by violent ideological groups that have an online presence?
- What issues, emotions, topics, and moral disengagement strategies in messaging foreshadow violent incidents?

We also conducted a series of controlled, randomized experiments to further examine phenomena that we observed (or anticipated) in the trace data. We attempted to answer the following questions:

- How do different types of identity-based posts in social media compare in their influence on attitudes, emotions, recall, credibility, and dissemination intentions?
- How do messaging coherence and amplification tactics influence attitudes, emotions, recall, credibility, and dissemination intentions?

Each research question is addressed in detail in the sections below.

Study of Trace Data

Description of Datasets

We compiled data of digital traces left by ideological groups - from website and social media communications by international and domestic ideological groups and records of violent events in which

groups (leaders and affiliated members) participated. A brief description of each dataset is provided below.

Biteback: This dataset is comprised of microblog posts from environmental activists on a platform called Biteback. The posts described activist activities including violent and criminal acts (e.g., animal liberation, vandalism, sabotage, and arson).

Immigration: This dataset is comprised of tweets containing keywords related to immigration (e.g., immigrant, alien) from the lead up to the 2016 presidential election and through inauguration.

Jihadi: This dataset contains tweets from jihadi militants and sympathizers prior to and following terrorist attacks by jihadi groups in Europe and the United States.

Domestic Ideological Groups (DIG): This dataset contains tweets from 70 extreme ideological groups, prominent group members, followers, and members of the public who interact with the groups. Groups were identified from lists maintained by the Southern Poverty Law Center's (SPLC) Hatewatch blog, a report on left-wing extremism prepared for the U.S. Department of Energy, the Counter Extremism Project (CEP), and The Armed Conflict Location & Event Data Project (ACLED).

Violent Events: This dataset contains details about violent events in which groups in the DIG dataset were involved. The dataset was compiled by searching newspaper articles for mentions of violent events and group names and manually verifying the groups' participation in each event.

Protection of Civil Liberties in Trace Data

Archival analyses performed in Year 2 examined publicly available datasets. Analysis of these datasets was approved by our Institutional Review Board (IRB). Additionally, we took steps to protect the identity of those whose posts are included in our datasets by encrypting all twitter sources and handles mentioned in tweets. Thus, analyses are performed on anonymized social media data. Additionally, no verbatim tweets or messages were (or will be) reproduced in published manuscripts or presentations.

How do violent and non-violent ideological groups use cognitive and affective processes in messaging?

We used the DIG dataset to answer this question and used correlational analyses to determine how violent and non-violent groups invoke cognitive and affective processes. We noted the following messaging differences in cognitive and affective processes across violent and non-violent groups:

- Violent ideological group messaging focuses on mostly negative emotional appeals (anger, anticipation, fear, and sadness) as well as on insight (e.g., think, know) and differentiation (e.g., hasn't, but, else) when referencing social identity only.

This finding suggests that violent ideological group leaders exert emotionally evocative media and fear appeals to foster social identity. In addition, they need to explain and justify the group's stances and emphasize differences with outgroups.

- Non-violent ideological group messaging focuses on mostly positive emotions (anticipation, surprise, joy, trust) when referencing issues. They also focus on positive emotions when referencing social identity, but they additionally rely on fear appeals, and all cognitive processes,

especially causality (e.g., because, effect), discrepancy (e.g., should, would) and certainty (e.g., always, never).

This finding suggests that non-violent ideological leaders focus on a wider array of issues and information, linking these to positive emotions to develop a sense of identity. Moreover, they foster social identity through a focus on group agency and future possibilities, or in other words, on the group's ability to have the power and resources to accomplish goals.

All correlations for language features are displayed below.

Cognitive and Affective Processes in Social Identity Messaging for Violent Groups

Table 1

Descriptive Statistics and Correlations among Variables for Violent Ideological Groups

|	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 Issues	4.65	2.23	-															
2 Social Identity	1.15	0.48	0.31	-														
3 Anger	0.36	0.15	0.21	0.50*	-													
4 Anticipation	0.39	0.10	-0.01	0.42*	0.69**	-												
5 Disgust	0.21	0.10	0.19	0.38	0.70**	0.61**	-											
6 Fear	0.42	0.17	0.23	0.61**	0.93**	0.65**	0.61**	-										
7 Joy	0.29	0.12	0.03	0.31	0.48*	0.82**	0.48*	0.35	-									
8 Sadness	0.28	0.13	0.28	0.64**	0.89**	0.72**	0.76**	0.93**	0.48	-								
9 Surprise	0.20	0.08	0.20	0.44*	0.76**	0.63**	0.74**	0.64**	0.67**	0.70**	-							
10 Trust	0.54	0.19	0.13	0.40	0.67**	0.78**	0.64**	0.65**	0.82**	0.72**	0.73**	-						
11 Insight	1.01	0.47	0.02	0.54**	0.01	0.18	0.29	0.10	0.14	0.33	0.11	0.09	-					
12 Causality	0.91	0.34	0.17	0.24	0.21	0.46*	0.39	0.30	0.30	0.43*	0.08	0.48*	0.33	-				
13 Discrepancy	0.76	0.32	0.37	0.50*	0.34	0.31	0.55**	0.23	0.30	0.38	0.45*	0.19	0.54**	0.24	-			
14 Tentativeness	1.17	0.73	0.21	0.17	0.02	0.01	0.11	0.00	0.11	0.09	0.21	0.17	0.20	0.47*	0.40	-		
15 Certainty	0.93	0.45	0.28	0.49*	0.19	0.43*	0.38	0.22	0.48*	0.39	0.20	0.48*	0.61**	0.62**	0.59**	0.40	-	
16 Differentiation	1.33	0.66	0.28	0.63**	0.44*	0.51*	0.65**	0.40	0.37	0.57**	0.36	0.40	0.53**	0.58**	0.74**	0.32	0.62**	-

*Note: N = 23. *p< .05. **p< .01.*

Cognitive and Affective Processes in Social Identity Messaging for Nonviolent Groups

Table 2

Descriptive Statistics and Correlations among Variables for Non-Violent Ideological Groups

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 Issues	5.31	1.76	-															
2 Social Identity	1.30	0.42	0.13	-														
3 Anger	0.38	0.14	0.07	0.28	-													
4 Anticipation	0.48	0.16	0.41*	0.44**	0.26	-												
5 Disgust	0.23	0.10	0.18	0.29	0.82**	0.29	-											
6 Fear	0.50	0.18	0.18	0.41*	0.88**	0.40*	0.71**	-										
7 Joy	0.37	0.13	0.34*	0.55**	0.26	0.89**	0.41*	0.46**	-									
8 Sadness	0.33	0.11	0.20	0.31	0.91**	0.40*	0.89**	0.87**	0.48*	-								
9 Surprise	0.24	0.06	0.37*	0.32	0.71**	0.59**	0.58**	0.76**	0.51**	0.76**	-							
10 Trust	0.73	0.22	0.48**	0.49**	0.22	0.77**	0.20	0.50**	0.75**	0.38*	0.63**	-						
11 Insight	1.19	0.30	-0.01	0.46**	0.15	0.29	0.18	0.18	0.32	0.23	0.26	0.36*	-					
12 Causality	1.01	0.27	0.04	0.49**	0.52**	0.28	0.55**	0.56**	0.45**	0.62**	0.43**	0.26	0.60**	-				
13 Discrepancy	0.87	0.33	0.23	0.59**	0.25	0.30	0.39*	0.23	0.39*	0.33*	0.25	0.15	0.28	0.59**	-			
14 Tentativeness	1.23	0.46	-0.20	0.43**	0.37*	0.27	0.36*	0.28	0.34*	0.40*	0.28	0.04	0.35*	0.62**	0.77**	-		
15 Certainty	1.01	0.37	-0.07	0.56**	0.25	-0.05	0.41*	0.22	0.19	0.30	0.18	0.05	0.24	0.41*	0.69**	0.57**	-	
16 Differentiation	1.44	0.51	-0.14	0.41*	0.54**	0.20	0.52**	0.46**	0.32*	0.57**	0.34*	0.03	0.34	0.71**	0.74**	0.89**	0.57**	-

*Note: N = 38. *p< .05. **p< .01.*

What dissemination tactics do violent ideological groups use prior to and following violent incidents?

To examine dissemination, we examined retweet rates in the DIG dataset and we calculated the difference in days between the date on which the tweet was posted and the collection date. The retweet rate was obtained by dividing the number of retweets by the number of days since the tweet was posted. These tweet-level retweet rates were averaged by group to obtain group-level retweet rates. Several of the groups showed a consistent retweeting pattern, however, there were some exceptions. The figure below shows the retweet rate of all groups which had a rate greater than zero over time. Since the retweet rates peaked in the latter years, the graph shows the rates for 2020 and 2021. As shown by the figure, the highest peaks in retweet rates were seen in early November and again in early January. Some of the events associated with these dates are the 2020 presidential election, the presidential inauguration, and the capitol riots on January 6th.

In the latter years, Retweet Rates Spiked Around the Election Time and in Early January
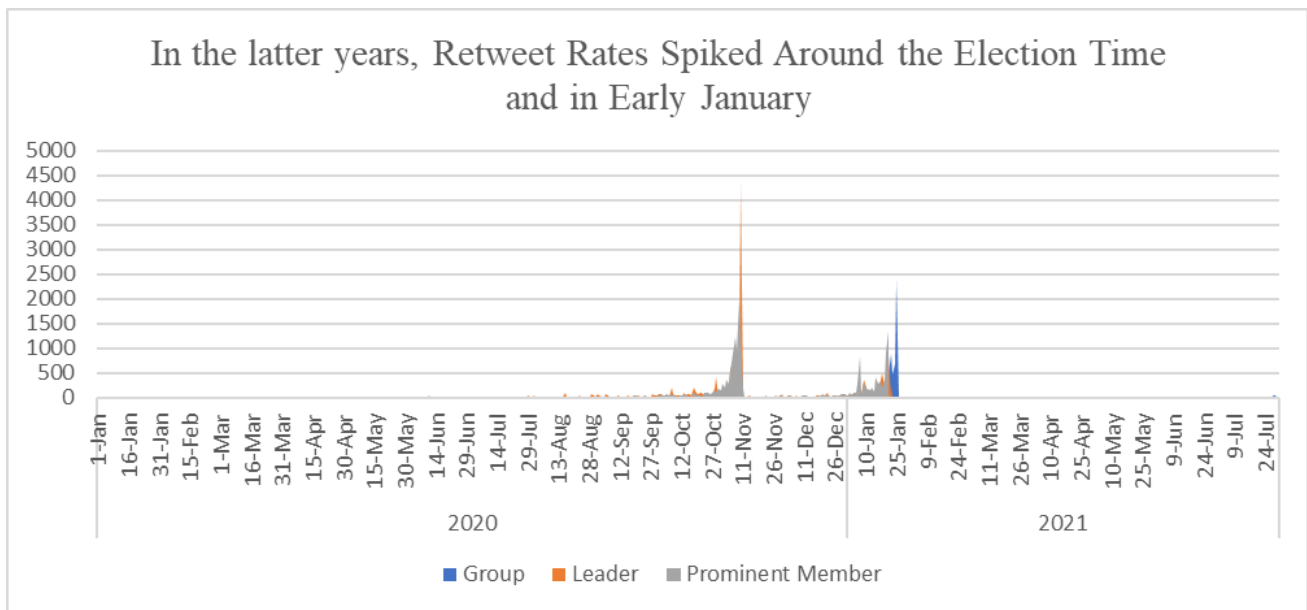
The figure below shows the retweet rate by group. The five most highly disseminated groups include: III%ers, American Renaissance, Traditionalist Worker Party, Center for Security Policy, and Blood & Honour.

Retweet Rate By Group
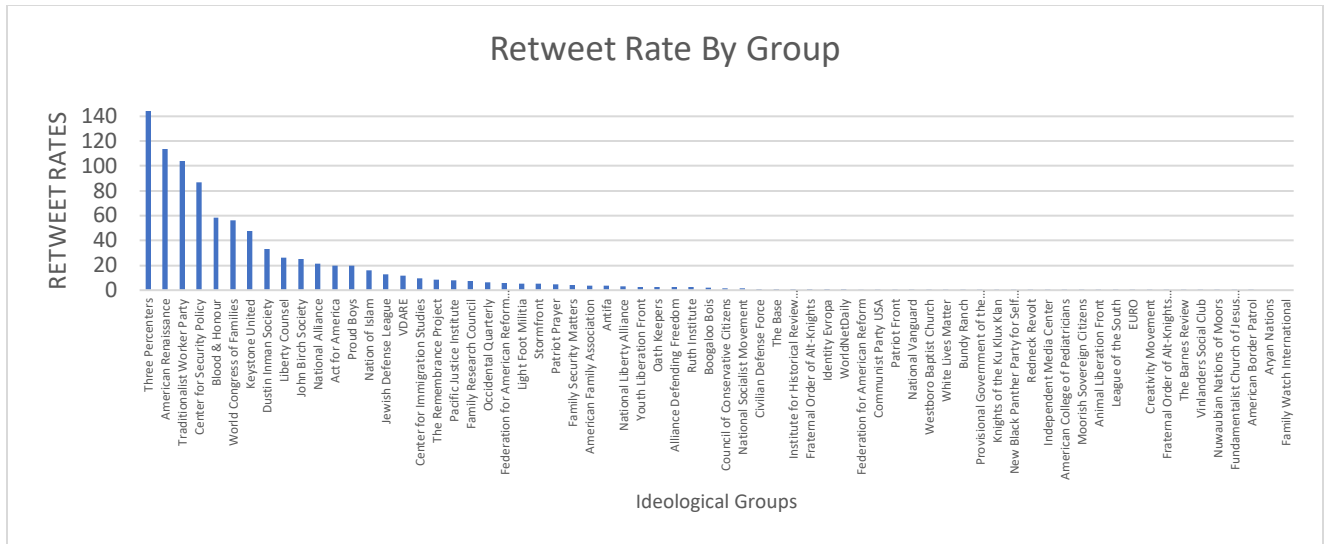
A closer qualitative look at the content of the tweets reveals that the context of a majority of tweets is related to political events, especially, events surrounding the government and presidency. The following figure shows the most prominent issues of interest across groups over time.

Issues that Spark the Most Interest Among Ideological Groups
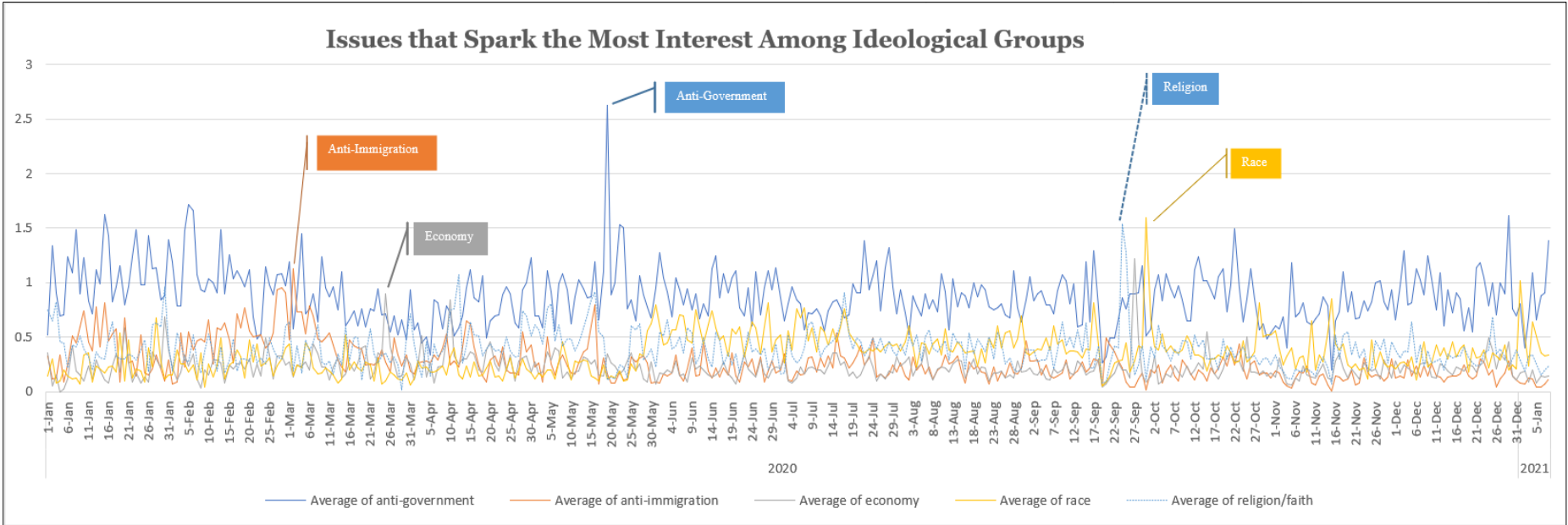
In summary, there is a remarkable sudden emergence of groups during contentious time periods like presidential elections and critical events.

There is a notable rapid spike of messages around particular events and then no spread at other times. Some scholars refer to slower inactive periods as "active abeyance", where group members strategically withdraw "from traditional public approaches to activism and recruitment. Instead, they engaged in more informal, private forms of activism directed at sustaining and growing the movement" Simi, P., & Futrell, R. (2020), p. 112.

The most prevalent topic is characterized by anti-government sentiment, followed by race-related topics. Religion and anti-immigration sentiments also peak at certain periods.

What justifications do violent and non-violent groups use for their actions?

Individuals consistently rely on 5 moral foundations (I.e., care, fairness, loyalty, authority, and sanctity) to judge and express whether certain actions and/or people are praiseworthy or blameworthy. Understanding what moral foundations violent and non-violent ideological groups mostly rely upon on social media can help us understand the lenses through which they judge the world. Running the LIWC moral foundations dictionary on 809,724 tweets from 66 ideological groups in the DIG dataset (26 violent and 40 non-violent) and comparing group means, we found the following patterns:

- Non-violent ideological group affiliates (I.e., group accounts, leaders, prominent members, followers and public) reference a larger proportion of positive fairness keywords (e.g., equality, justice, rights).
- Non-violent ideological public (I.e., individuals who mention the group's name or share their content on social media but do not follow group accounts) reference a larger proportion of positive care (e.g., kindness, compassion, empathy) and negative fairness keywords (e.g., cheat, fraud, unfair, injustice).
- Violent ideological group affiliates reference a larger proportion of positive sanctity keywords (e.g., purity, sacred, wholesome).
- Violent group leaders reference a larger proportion of negative authority keywords (e.g., subversion, disobey, disrespect, chaos).

What are the types and timings of crimes committed by violent online ideological groups?

Of the 34 groups that had committed at least one violent act, nine were each associated with a single violent event between January 2016 and December 2021, 18 were associated with between 2 and 10 events, and 3 were associated with between 11 and 20 events. Antifa was associated with 22 violent events and the White Lives Matter group was associated with 23 violent events.

We are uncovering distinct profiles of violent groups based on the type of acts committed by groups, i.e., property damage, cyber assault (e.g., doxxing, swatting), intimidation, injuries (single vs multiple), and deaths (single vs multiple). The summary of violent actions committed by group members is shown below.
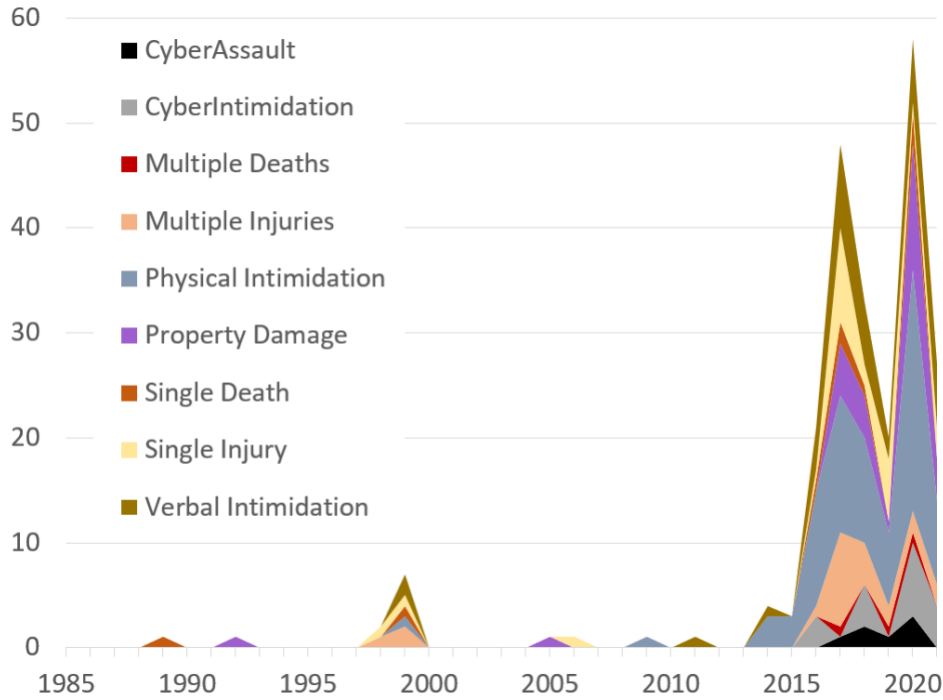
What issues, emotions, topics, and moral disengagement strategies in messaging foreshadow violent incidents?

For the DIG groups participating in the Jan 6 Capitol riot, we analyzed their tweets in the 8-week period around the riots. We found a rise in a topic evincing religious rhetoric with low emotiveness foreshadowing the Jan 6 event. This rise in religious rhetoric is consistent with our earlier findings on the dataset of Jihadi tweets around violent events.

Foreshadowing Violence

- Analyzed corpus of 3,615 tweets from domestic ideological groups that participated in the January 6th Capitol riot (OathKeepers, ProudBoys, and IIIPercenters)
- Performed text mining (structural topic modeling) on tweets within an 8-week window around the focal event (4 weeks before and 4 weeks after)
- Selected a 10-topic model selected to optimize coherence and exclusivity
- **Topic 1** was **most prevalent** and revealed use of **religious rhetoric**
- High frequency and exclusivity **religious words**: god and evil
- **Topic 1 foreshadowed the focal event**, occurring more frequently in tweets before than after the event (t=2.57)
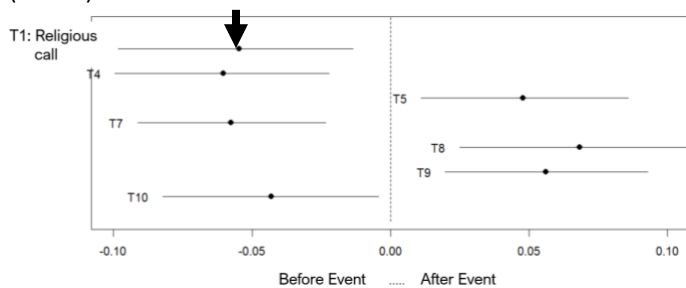
- These **tweets were dispassionate**: tweets in which this topic dominated reflected the lowest negative emotion and the second lowest positive emotion (range of t=4.36 for disgust to t=11.01 for trust)
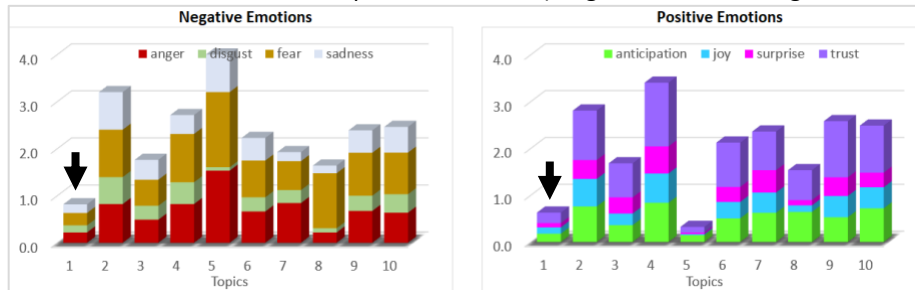
Experimental Studies and Findings

How do identity-based posts in social media compare in their influence on attitudes, emotions, recall, credibility, and dissemination intentions?

Voter ID and Social Identity Study – Minority and Political Social Identities

This randomized controlled online experiment examined how the presence of social identity information in an ideological twitter feed about voter identification requirements influenced attitudes towards the topic, emotional reactions, recall of specific tweets, credibility of the tweets, and dissemination intentions (like, retweet, hashtag, share). A simulated Twitter feed was created that included 1 initiating anti-voter ID tweet and 12 reactionary tweets reflecting pro- and anti-voter ID perspectives. Social identity information was manipulated across 4 unique study conditions, with the tweets otherwise remaining exactly the same. Condition one contained social identity information related to racial minorities in 8 tweets. Condition two contained social identity information related to political orientations (e.g., conservatives, liberals, democrats, republicans) in 8 tweets. Condition three contained both minority and political social identity information. Condition four contained no explicit social identity information. A battery of covariate measures included digital activism, political orientation, general social media usage, social desirability, and the big-5 personality variables. Coefficient alpha scale reliabilities ranged from .70 to .88. A series of logistic regressions and Analysis of Covariance analyses were conducted to examine main and interactive effects of minority and political social identity content on the outcomes of interest. Covariates were only included in an analysis if they were significant. For simplicity, covariate results are not included in the summary below.

Attitudes

Attitudes about the topic were assessed before and after participant exposure to the tweets. They indicated how pro-voter ID they were on a five point scale (stance) and responded to several questions regarding strength of their attitude and personal importance of the issue.

Across all four conditions, exposure to this social media exchange resulted in increases in participants feeling more strongly about the topic ($t = -3.00$, $p < .05$) and moved them towards an anti-voter ID stance ($t = -2.56$, $p < .05$). However, exposure to minority identity, political identity, or the combination of both social identities did not significantly change stance or strength of stance after exposure to the tweet feed.

Emotional Reactions

Emotional reactions to the feed were evaluated using linguistic coding of participants' written tweet response to the feed. The coding scheme used Plutchik's (1980) emotions to assess negative and positive emotions.

Participants in conditions that included political identity content expressed less negative emotion including anger, sadness, disapproval, and aggression in their response to the tweet than those who did not see political identity content (negative emotion $F = 12.32$, $p < .001$, $\eta p2 = .06$; anger $F = 13.30$, $p < .001$, $\eta p2 = .06$; sadness $F = 4.49$, $p < .05$, $\eta p2 = .02$, disapproval $F = 8.32$, $p < .01$, $\eta p2 = .04$; aggression $F = 5.86$, $p < .05$, $\eta p2 = .03$). Interestingly, participants exposed to minority identity content showed significantly more aggression in their tweet responses than those not exposed to minority identity content ($F = 5.96$, $p < .05$, $\eta p2 = .03$). Minority identity did not show any other main effects for emotion.

Recall & Mentioning

Participants were asked to restate as many of the anti-voter ID tweets as they could remember and to restate as many of the pro-voter ID tweets they could remember. A team of three trained coders rated on 3-point scale the extent to which information from each tweet in the feed was reflected in these restatements. These were aggregated to create anti and pro-voter ID recall scores. Participants' response to the tweet feed was also evaluated for the extent to which it mentioned anti-voter ID tweets and the extent to which it mentioned pro-voter ID tweets.

People exposed to minority identity recalled significantly more anti-voter ID tweets than those not exposed ($F = 17.35$, $p < .001$, $\eta p2 = .08$). Additionally, minority and political identity significantly interacted to influence anti-voter ID tweets recall ($F = 20.86$, $p < .001$, $\eta p2 = .10$) such that those exposed to minority identity content recalled the most while those exposed to neither minority or political identity content recalled the least. Similar results were seen for pro-voter ID tweet recall with participants exposed to minority identity content (vs. not) recalling more pro-voter ID tweets ($F = 9.65$, $p < .01$, $\eta p2 = .05$). Those exposed to minority identity content recalled the most pro-voter ID tweets, while those exposed to neither minority or political identity content recalled the least ($F = 26.12$, $p < .001$, $\eta p2 = .12$).

With regard to mentioning, minority and political social identity content both exerted main effects, with more mentioning of anti-voter ID tweets when each of these identities was present (minority $F = 7.52$, $p < .01$, $\eta p2 = .03$, political $F = 4.46$, $p < .05$, $\eta p2 = .02$). The interaction was also significant with the minority identity exposure resulting in the most mentions of anti-voter ID tweets and no social identity exposure resulting in the least mention of these ($F = 5.98$, $p < .05$, $\eta p2 = .03$). This interaction was also significant for mention of pro-voter ID tweets ($F = 10.64$, $p = .001$, $\eta p2 = .05$).

Credibility
Participants rated message credibility of each tweet by rating the trustworthiness, fairness, expertise, goodwill, and currency of the tweet content. These were aggregated to create anti- and pro-voter ID credibility scores.

Exposure to social identity content did not significantly affect credibility evaluations of pro-voter ID or anti-voter ID tweets.

Dissemination Intentions
Participants were asked to rate the extent to which they would like (Y/N), hashtag words in the initial anti-voter ID tweet (Y/N), a pro-voter ID tweet and an anti-voter ID tweet. They were also asked whether they would retweet an anti-voter ID and a pro-voter ID tweet and whether they would share the feed with similar others (people who share the participant's views on the issue) and with dissimilar others (people with different views on the issue). The like and hashtag outcomes were analyzed using logistic regressions given their dichotomous nature. The sharing scores ranged from 0 to 2 for the initial, anti- and pro-voter ID retweets and were analyzed using ANCOVAs.

Logistic regressions
Initial tweet results. Participants exposed to minority identity content were 4.46 times more likely to like the initial anti-voter ID tweet, which is anti-voter ID ($b = 1.50$, $p < .05$). Participants exposed to both social identities were significantly less likely to like this tweet ($b = -1.87$, $p < .05$). Social identity content exposure was not related to hashtaging words in the initial tweet.

Pro-voter ID tweet results. Participants exposed to political identity were .22 times more likely to like the pro-voter ID tweet (b = 1.96, p<.01).

Anti-voter ID tweet results. Exposure to social identity content was not a significant predictor of anti-voter ID tweet liking.

ANCOVAs
Retweet results. Exposure to social identity content was not a significant predictor of the initial, pro-voter ID, or anti-voter ID tweet retweet.

Sharing entire feed with others results. Exposure to social identity content was not a significant predictor of sharing the social media feed with similar or dissimilar others.

Sharing initial, pro-voter, and anti-voter ID tweets with others results. Exposure to social identity content was not a significant predictor of sharing the initial or the pro-voter ID tweet with similar or dissimilar others. Nor did social identity influence sharing the anti-voter ID tweet with similar others. However, minority and political identity interacted to influence sharing of the anti-voter ID tweet with dissimilar others (F = 4.39, p < .05, ηp2 = .02). Those exposed to minority identity and no political identity had the highest intention to share anti-voter ID tweets with dissimilar others.

Moderated Mediation Analysis
We also wanted to investigate the possibility that the moderated relationship between minority social identity (moderator = political identity) and information recall could be mediated by negative emotions. Emotions have been shown to direct attention and memory processes in prior research. Results of a series of moderated mediation PROCESS models run in SPSS showed some interesting findings. First, the influence of minority identity content on the recall of anti-voter ID tweets was mediated by aggression, but this effect was conditional. Interestingly, when political social identity content was present along with minority identity content, the relationship between minority identity and aggression was suppressed and the relationship between aggression and the recall of anti-voter ID tweets was suppressed. Political identity content also moderated the direct relationship of minority identity content to anti-voter ID recall. These findings are shown in Figure 1 and Table 1 below.
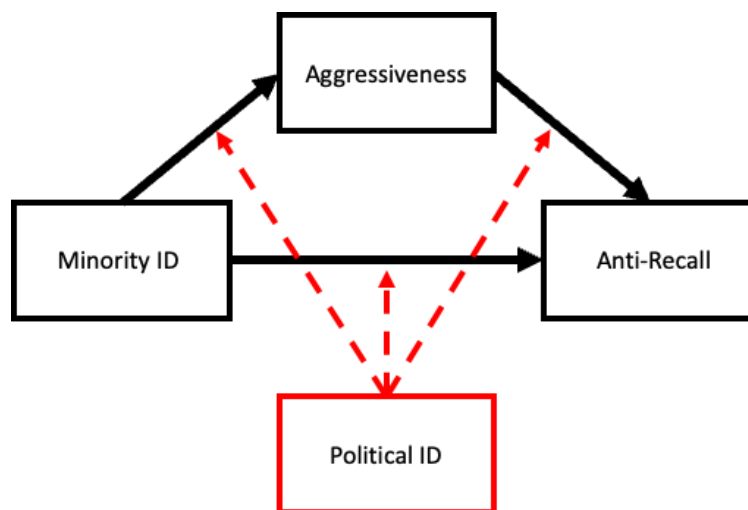
Figure 1.

Table 1.

Moderated Mediation Analysis: Anti Recall Predicted by Minority Identity Mediated by Aggressiveness and Moderated by Political Identity

Variable	Estimate	SE	95% CI		p
			LL	UL	
Mediator Model					
Intercept	.6000	.1752	.2544	.9456	.0008
Minority Identity (X)	.5455	.2421	.0679	1.023	.0254
Political Identity (W)	-.3021	.2518	-.7986	.1944	.2316
Minority Identity*Political Identity(XW)	-.2515	.3502	-.9421	.4391	.4735
Dependent Variable Model					
Intercept	.9593	.2065	.5519	1.367	.0000
Minority Identity (X)	1.5902	.2789	1.040	2.140	.0000
Aggressiveness (M)	.2346	.0965	.0442	.4250	.0160
Political Identity (W)	.9158	.2939	.3362	1.495	.0021
Minority Identity*Political Identity(XW)	-1.641	.4019	-2.4334	-.8480	.0001
Aggressiveness*Political Identity (MW)	-.3865	.1754	-.7323	-.0406	.0287
	Effect	SE	LLCI	ULCI	p
Conditional Direct Effect of X on Y Political Identity Condition					
0	1.590	.2789	1.040	2.140	.0000
1	-.0504	.2894	-.6212	.5203	.8618
	Indirect Effect	Boot SE	Boot LLCI	Boot ULCI	
Conditional Indirect Effects at values of the moderator Political Identity Condition					
0	.1279	.0851	-.0086	.3215	
1	-.0446	.0468	-.1571	.0301	
	Index	Boot SE	Boot LLCI	Boot ULCI	
Index of moderated mediation					
Political Identity	-.1726	.0970	-.3795	-.0091	

A second key finding from these PROCESS model analyses is that the relationship between minority identity and sharing anti-voter ID tweets with dissimilar others was also mediated by the amount of anti-voter ID tweets recalled, but this effect was conditional. Political identity content moderated (and suppressed) the relationship between minority identity content and anti-voter ID recall and between

anti-voter ID recall and dissemination to dissimilar others. Thus, political identity appears to have interfered with the digital activism patterns of emotion and recall and of recall and dissemination that were observed when only minority identity was present. These findings are shown in Figure 2 and Table 2 below.
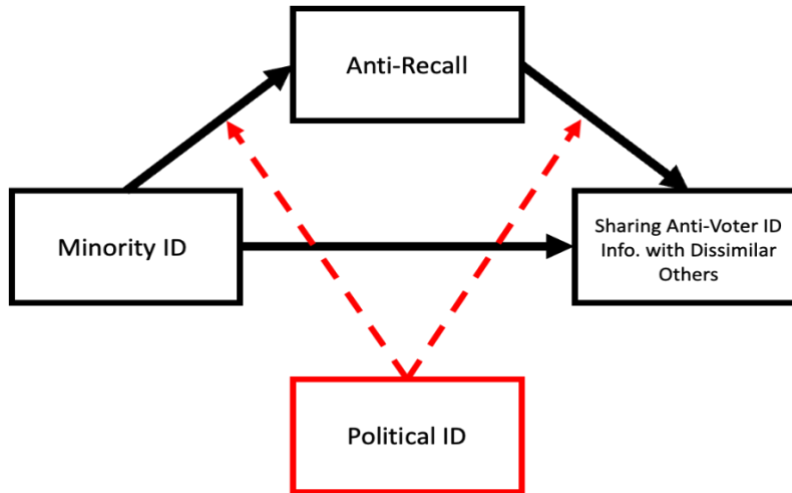
Figure 2.

Table 2.

Moderated Mediation Analysis: Sharing Anti Voter ID Information with Dissimilar Others Predicted by Minority Identity Mediated by Anti Recall and Moderated by Political Identity

Variable	Estimate	*SE*	95% CI		*p*
			LL	*UL*	
Mediator Model					
Intercept	1.100	.2007	.7041	1.496	.0000
Minority Identity (X)	1.7182	.2773	1.171	2.265	.0000
Political Identity (W)	.7298	.2884	.1611	1.299	.0122
Minority Identity*Political Identity(XW)	-1.813	.4011	-2.604	-1.022	.0000
Dependent Variable Model					
Intercept	1.2541	.1343	.9891	1.519	.0000
Minority Identity (X)	-.0518	.1315	-.3111	.2074	.6937
Anti-Recall (M)	.1723	.0539	.0650	.2786	.0016
Political Identity (W)	.1928	.2004	-.2024	.5880	.3372
Anti-Recall*Political Identity (MW)	-.1743	.0850	-.3420	-.0067	.0416
Direct Effect of X on Y	-.0518	.1315	-.3111	.2075	.6937

Conditional Indirect Effects at values of the moderator	*Indirect Effect*	*Boot SE*	*Boot LLCI*	*Boot ULCI*
Political Identity Condition				
0	.2960	.1324	.0587	.5794
1	.0002	.0150	-.0394	.0251

Index of moderated mediation	*Index*	*Boot SE*	*Boot LLCI*	*Boot ULCI*
Political Identity	-.2958	.1328	-.5845	-.0612

Death Penalty Social Media Study - Minority and Pro-life Social Identities

This randomized controlled online experiment examined how the presence of social identity information in an ideological twitter feed about the death penalty influenced attitudes towards the topic, emotional reactions, recall of specific tweets, credibility of the tweets, and dissemination intentions (like, retweet, hashtag, share). A simulated Twitter feed was created that included 1 initiating anti-death penalty ID tweet and 12 reactionary tweets reflecting pro- and anti-death penalty perspectives. Social identity information was manipulated across 4 unique study conditions, with the tweets otherwise remaining exactly the same. Condition one contained social identity information related to racial minorities in 8 tweets. Condition two contained social identity information related to prolife identity in 8 tweets. Condition three contained both minority and prolife social identity information. Condition four contained no explicit social identity information. A battery of covariate measures included digital activism, political orientation, general social media usage, and the big-5 personality variables. Coefficient alpha scale reliabilities ranged from .72 to .88. The social desirability scale was dropped due to low reliability. A series of logistic regressions and Analysis of Covariance analyses were conducted to examine main and interactive effects of minority and pro-life social identity content on the outcomes of interest. Covariates were only included in an analysis if they were significant. For simplicity, covariate results are not included in the summary below.

Attitudes

Attitudes about the topic were assessed before and after participant exposure to the tweets. They indicated how pro-death penalty they were on a five point scale (stance) and responded to several questions regarding strength of their attitude and personal importance of the issue.

Across all four conditions, exposure to this social media exchange resulted in increases in participants feeling more strongly about the topic and moved towards an anti-death penalty stance, but the issue was less personally important to them after exposure to the feed. However, exposure to minority identity, pro-life identity, or the combination of both social identities did not significantly change stance or strength of stance after exposure to the tweet feed.

Emotional Reactions

Emotional reactions to the feed were evaluated using linguistic coding of participants' written tweet response to the feed. The coding scheme used Plutchik's (1980) emotions to assess negative and positive emotions.

Negative emotions overall were not affected by exposure to social identity content, nor were the emotions of sadness, fear, or disapproval. Exposure to prolife identity content showed less anger than those who were not exposed to prolife content (F = 4.44, p <.05, ηp2 = .02). While this result initially seemed counter-intuitive, many individuals who hold pro-life views also tend to have conservative views that are consistent with supporting the death penalty. A similar pattern was seen for aggression (F = 4.44, p <.05, ηp2 = .02) where those exposed to pro-life identity content expressed less aggression in their tweet response than those not shown pro-life content.

Positive emotions overall were also not affected by social identity content exposure, however, participants exposed to pro-life content expressed less submission than those not shown this content.

Recall & Mentioning

Participants were asked to restate as many of the anti-death penalty tweets as they could remember and to restate as many of the pro-death penalty tweets they could remember. A team of three trained

coders rated on 3-point scale the extent to which information from each tweet in the feed was reflected in these restatements. These were aggregated to create anti and pro-death penalty recall scores. Participants' response to the tweet feed was also evaluated for the extent to which it mentioned anti-death penalty tweets and the extent to which it mentioned pro-death penalty tweets.

Prolife identity and minority identity content significantly interacted to influence the recall of all the tweets ($F = 39.98$, $p < .001$, $\eta p2 = .16$). Those exposed to minority identity content but not to prolife identity content recalled the most, followed by those exposed to prolife identity content but not to minority identity content. Exposure to either identity, but not both, led to better recall of tweets. This same interaction pattern held for recall of pro death penalty tweets ($F = 31.09$, $p < .001$, $\eta p2 = .13$).

Exposure to prolife identity significantly affected the recall of anti-death penalty tweets ($F = 5.18$, $p < .05$, $\eta p2 = .02$). The interaction between prolife identity and minority identity content significantly affected the recall of anti-death penalty tweets ($F = 33.35$, $p < .001$, $\eta p2 = .14$). Those exposed to minority identity content but not to prolife identity content recalled the most pro death penalty tweets, followed by those exposed to prolife identity content but not to minority identity content. It seems that being exposed to either identity led to better recall of anti-death penalty tweets.

Exposure to minority identity content ($F = 5.2$, $p < .05$, $\eta p2 = .02$) led to greater mention of tweets (pro and anti) in the tweet response than no exposure to minority content while the opposite was true for exposure to prolife identity content ($F = 26.71$, $p < .001$, $\eta p2 = .12$). Prolife identity and minority identity content also significantly interacted to affect the mention of all tweets ($F = 23.66$, $p < .001$, $\eta p2 = .10$). Those exposed to minority identity content but not to prolife identity content mentioned a larger extent of content from all tweets in their responses, followed by those not exposed to either identity. Those exposed to both identities mentioned the least amount of content from all the tweets.

Exposure to prolife identity content ($F = 5.65$, $p < .05$, $\eta p2 = .03$) significantly reduced mentions of pro death penalty tweets in their tweet response. The interaction between prolife identity and minority identity content also significantly affected the mention of pro death penalty tweets ($F = 20.30$, $p < .001$, $\eta p2 = .07$). Those exposed to minority identity content but not to prolife identity content mentioned a larger extent of content from pro death penalty tweets in their responses, followed by those exposed to prolife identity but not minority identity. It seems that being exposed to either identity alone led to more mention of pro death penalty tweets. Those exposed to both identities mentioned the least amount of content from pro death penalty tweets.

Exposure to prolife identity content ($F = 26.66$, $p < .001$, $\eta p2 = .12$) significantly affected the extent to which participants mentioned content from anti-death penalty tweets in their responses to the tweets. The interaction between prolife identity and minority identity content significantly affected the mention of anti-death penalty tweets, too ($F = 9.94$, $p < .01$, $\eta p2 = .05$). Those exposed to minority identity content but not to prolife identity content mentioned a larger extent of content from anti-death penalty tweets in their responses, followed by those not exposed to either identity. Those exposed to both identities mentioned the least amount of content from anti-death penalty tweets.

Credibility
Participants rated message credibility of each tweet by rating the trustworthiness, fairness, expertise, goodwill, and currency of the tweet content. These were aggregated to create anti- and pro-death penalty credibility scores.

Exposure to social identity content did not significantly affect credibility evaluations of pro- or anti-death penalty tweets.

Dissemination Intentions
Participants were asked to rate the extent to which they would like (Y/N) and hashtag words in the initial anti-death penalty tweet (Y/N), a pro-death penalty tweet, and an anti-death penalty tweet. They were also asked whether they would retweet an anti- death penalty and a pro-death penalty tweet and whether they would share the feed with similar others (people who share the participant's views on the issue) and with dissimilar others (people with different views on the issue). The liking and hashtag outcomes were analyzed using logistic regressions given their dichotomous nature. The sharing scores ranged from 0 to 2 for the initial, anti- and pro-death penalty tweets and were analyzed using ANCOVAs.

Logistic regressions
Initial tweet results. Participants exposed to prolife identity content were .348 less likely to like the initial anti-death penalty tweet (b = -1.05, p < .05). Social identity content exposure was not related to hashtaging words in the initial tweet.

Pro-death penalty tweet results. Participants exposed to prolife identity content (b = .97, p < .05) were 2.6 times more likely to like the pro death penalty tweet. However, social identity content did not influence hashtaging.

Anti-death penalty tweet results. Exposure to social identity content was not a significant predictor of anti-death penalty tweet liking.

ANCOVAs
Retweet results. Exposure to social identity content was not a significant predictor of retweeting the initial pro-death penalty tweet or the other pro-death penalty tweet. However, prolife identity and minority identity content interacted to influence retweet of the anti-death penalty tweet (F = 4.08, p < .05, ηp2 = .02) Participants exposed to both minority identity and prolife identity content had higher intentions to retweet the anti-death penalty tweet, whereas those only exposed to prolife identity content had the lowest intentions to retweet it.

Sharing entire feed with others results. Exposure to social identity content was not a significant predictor of sharing the social media feed with similar or dissimilar others.

Sharing initial, pro-death penalty, and anti-death penalty tweets with others results. Exposure to social identity content was not a significant predictor of sharing the initial or the pro-death penalty tweets with similar or dissimilar others. Nor did social identity influence sharing the anti-death penalty tweet with similar others. However, minority and political identity interacted to influence sharing of the anti-voter ID tweet with dissimilar others (F = 4.39, p < .05, ηp2 = .02). Those exposed to minority identity and no political identity had the highest intention to share anti-voter ID tweets with dissimilar others. Exposure to minority identity content (F = 3.97, p < .05, ηp2 = .02) resulted in greater intentions to disseminate death penalty tweets overall.

Mediated Moderation Analyses
We also wanted to investigate the possibility that the moderated relationship between pro-life identity (moderator = minority identity) and information recall could be mediated by negative emotions. Emotions have been shown to direct attention and memory processes in prior research. Results of a

series of moderated mediation PROCESS models run in SPSS showed some interesting findings. First, the influence of pro-life identity content on the recall of all tweets was mediated by aggression, but this effect was conditional and depended on minority identity also being present. Interestingly, when minority social identity content was present along with pro-life identity content, the relationship between pro-life identity and aggression was suppressed and the relationship between aggression and the recall of all tweets was suppressed. These findings are shown in Figure 3 and Table 3 below.
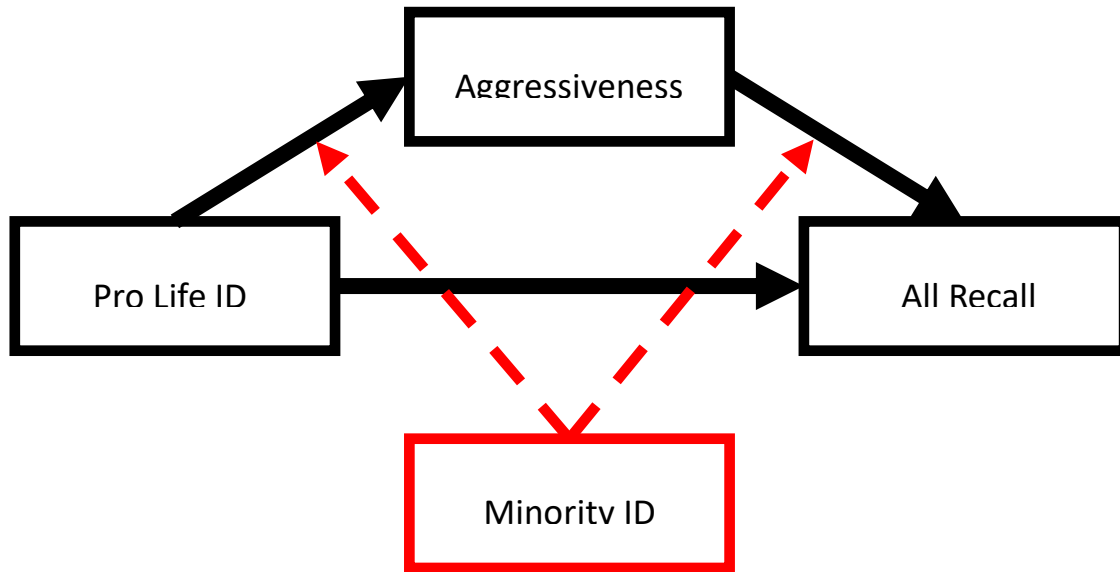
Figure 3.

Table 3.

Moderated Mediation Analysis: All Recall Predicted by Pro Life Identity Mediated by Aggressiveness and Moderated by Minority Identity

Variable	Estimate	SE	95% CI		p
			LL	UL	
Mediator Model					
Intercept	3.926	.5745	2.793	5.059	.0000
Pro Life Identity (X)	-.6852	.8125	-2.287	.9167	.4000
Minority Identity (W)	.3704	.8125	-1.2315	1.9723	.6490
Pro Life Identity*Minority Identity(XW)	-1.257	1.167	-3.558	1.044	.2827
Dependent Variable Model					
Intercept	4.0985	.2925	3.522	4.675	.0000
Pro Life Identity (X)	-.0930	.2825	-.6500	.4640	.7424
Aggressiveness (M)	.0061	.0445	-.0817	.0940	.1380
Minority Identity (W)	.0765	.3601	-.6335	.7866	.8319
Aggressiveness*Minority Identity (MW)	.1415	.0660	.0114	.2716	.03332
	-.0930	.2825	-.6500	.4640	.7424
Direct Effect of X on Y					
Conditional Indirect Effects at values of the moderator	Indirect Effect	Boot SE	Boot LLCI	Boot ULCI	
Minority Identity Condition					
0	-.0042	.0403	-.0541	.1196	
1	-.2867	.2177	-.8175	-.0118	
Index of moderated mediation	Index	Boot SE	Boot LLCI	Boot ULCI	
Minority Identity	-.2825	.2212	-.8274	-.0086	

A similar pattern of results occurred when looking at the relationship of pro-life identity content to anti-death penalty recall, mediated by aggression and moderated by minority identity content.
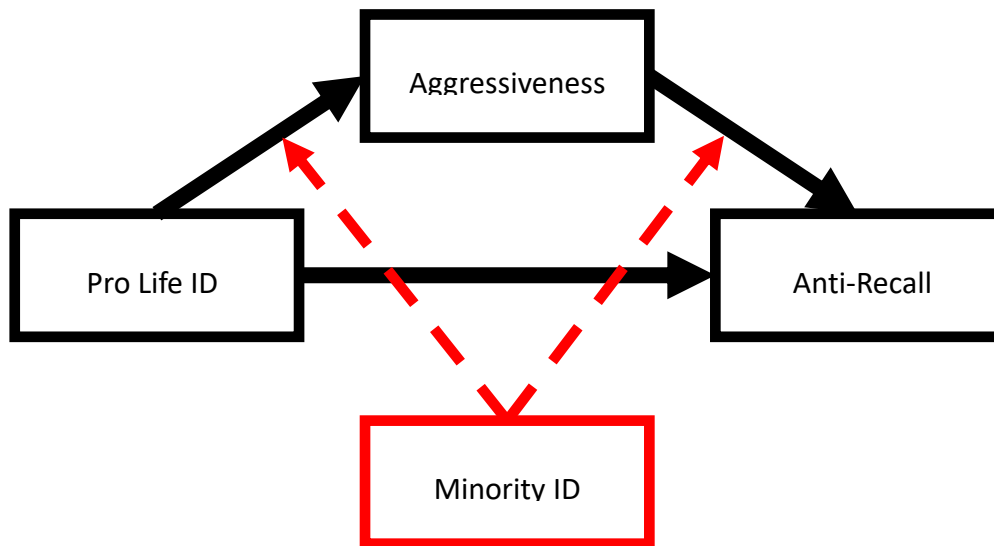
Figure 4.

Table 4.

Moderated Mediation Analysis: Anti Recall Predicted by Pro Life Identity Mediated by Aggressiveness and Moderated by Minority Identity

Variable	Estimate	SE	95% CI		p
			LL	UL	
Mediator Model					
Intercept	3.926	.5745	2.793	5.059	.0000
Pro Life Identity (X)	-.6852	.8125	-2.287	.9167	.4000
Minority Identity (W)	.3704	.8125	-1.2315	1.9723	.6490
Pro Life Identity*Minority Identity(XW)	-1.257	1.167	-3.558	1.044	.2827
Dependent Variable Model					
Intercept	2.1739	.1755	1.828	2.520	.0000
Pro Life Identity (X)	-.2029	.1695	-.5371	.1313	.2327
Aggressiveness (M)	.0004	.0267	-.0522	.0531	.9869
Minority Identity (W)	-.1134	.2161	-.5394	.3126	.6003
Aggressiveness*Minority Identity (MW)	.0997	.0396	.0217	.1778	.0125
Direct Effect of X on Y	-.2029	.1695	-.5371	.1313	.2327

Conditional Indirect Effects at values of the moderator	Indirect Effect	Boot SE	Boot LLCI	Boot ULCI	
Minority Identity Condition					
0	-.0003	.0237	-.0352	.0616	
1	-.1946	.1300	-.5070	-.0169	

Index of moderated mediation	Index	Boot SE	Boot LLCI	Boot ULCI
Minority Identity	-.1943	.1309	-.5074	-.0161

How do messaging coherence and application tactics influence on attitudes, emotions, recall, credibility, and dissemination intentions?

Gun Rights Study – Coherence, Credibility, and Amplification

This study examined the influences of message coherence, credibility, and amplification in a simulated feed from a fictitious ideological group that supports the second amendment. Participants were shown a simulated web page of the group that gave basic information about the groups purpose and then showed 12 tweets from a social media feed linked to this group. Outcomes included attitudes towards gun rights, emotional reactions, recall of specific tweets, credibility of the tweets, and dissemination intentions (like, retweet, hashtag, and share). A simulated Twitter feed was created that included 12 tweets from the official group account. Information in these tweets was manipulated to create high message coherence (all tweets were about gun rights) and low message coherence (4 tweets were about gun rights, 8 tweets covered other topics including immigration, climate change, capital punishment, separation of church and state, and voter ID). Amplification was also manipulated to be high (large numbers of likes and retweets for each tweet) or low (small numbers of likes and retweets). Credibility was manipulated to be high (accounts of tweeters show a verified symbol) or low (accounts of tweeters do not have a verified symbol

A battery of covariate measures included digital activism, political orientation, general social media usage, social desirability, and the big-5 personality variables. Coefficient alpha scale reliabilities ranged from .71 to .90. A series of logistic regressions and Analysis of Covariance analyses were conducted to examine main and interactive effects of coherence, amplification, and credibility on the outcomes of interest. Covariates were only included in an analysis if they were significant. For simplicity, covariate results are not included in the summary below.

Attitudes

Attitudes about the gun rights topic were assessed before and after participant exposure to the tweets. They indicated how pro-gun rights (vs. gun control) they were on a five point scale (stance) and responded to several questions regarding strength of their attitude and personal importance of the issue.

Coherence, amplification, credibility did not reveal any main effects or interactive effects on change in self-reported stance overtime or on change in importance from pre-feed to post-feed exposure.

Amplification positively predicted post feed strength of stance on the gun rights issue as reflected in their response tweet to the feed ($F = 3.68$, $p = .057$, $\eta p2 = .02$) with high amplification resulting in higher strength scores compared to low amplification. The interaction between coherence and amplification significantly predicted strength of stance on the gun rights issue ($F = 3.98$, $p < .05$, $\eta p2 = .02$). Those exposed to coherent gun-related tweets that were highly amplified expressed greater pro-gun rights attitudes than the rest. Those exposed to coherent gun-related tweets that were not amplified expressed the lowest pro-gun rights attitudes.

Coherence, amplification, credibility did not reveal any main or interactive effects on self- reported certainty of stance based on their feed response. The interaction between coherence, amplification and verification predicted certainty based on their response to feed ($F = 3.83$, $p = .052$, $\eta p2 = .02$). Those exposed to coherent, non-verified, non-amplified tweets expressed the highest certainty on their stance. Those exposed to coherent, non-verified, amplified tweets expressed the least certainty on their stance.

Coherence significantly predicted importance based on their feed response ($F = 7.26$, $p < .05$, $\eta p2 = .03$). Those exposed to coherent gun-related tweets expressed that the issue was more important to them than those who were exposed to non-coherent content.

Coherence also significantly predicted stance based on their feed response ($F = 4.33$, $p < .05$, $\eta p2 = .02$). Those exposed to coherent gun-related tweets expressed greater pro-gun rights attitudes than those who were exposed to non-coherent content.

Emotional Reactions
Emotional reactions to the feed were evaluated using linguistic coding of participants' written tweet response to the feed. The coding scheme used Plutchik's (1980) emotions to assess negative and positive emotions.

The interaction between amplification and credibility significantly predicted submission in their feed response ($F = 3.75$, $p = .054$, $\eta p2 = .02$). Those exposed to verified, non-amplified tweets expressed the highest submission in their response whereas those exposed to verified, amplified tweets expressed the least submission in their response. Amplification significantly predicted disapproval in their feed response ($F = 3.81$, $p = .052$, $\eta p2 = .02$). Those exposed to amplified tweets expressed more disapproval in their response than those exposed to non-amplified tweets. No other emotion scores were influenced by coherence, amplification, or credibility.

Recall & Mentioning
Participants were asked to restate as many of the gun control tweets as they could remember and to restate as many of the pro-gun rights tweets they could remember. A team of three trained coders rated on 3-point scale the extent to which information from each tweet in the feed was reflected in these restatements. These were aggregated to create gun control and pro-gun rights recall scores. Participants' response to the tweet feed was also evaluated for the extent to which it mentioned gun control tweets and the extent to which it mentioned pro-gun rights tweets.

There were no main or interactive effects for coherence, amplification, or credibility on recall of the feed as a whole. However, coherence significantly predicted recall of gun-related tweets (anti-gun control) ($F = 7.41$, $p < .01$, $\eta p2 = .04$) such that participants exposed to the low coherence feed (mix of gun-related and non-gun-related tweets) recalled more content from gun-related tweets than those exposed to the high coherence feed.

Credibility
Participants rated message credibility of each tweet by rating the trustworthiness, fairness, expertise, goodwill, and currency of the tweet content. These were aggregated to create anti-gun rights and pro-gun rights credibility scores. Participants exposed to the high coherence conditions saw the tweets as more credible than those in the low coherence condition ($F = 17.79$, $p < .001$, $\eta p2 = .09$) and they saw the ideological group as more credible ($F = 16.31$, $p < .001$, $\eta p2 = .08$) than those in the low coherence condition. There were no other significant main effects or interactions for credibility.

Dissemination Intentions
Analyses for dissemination intentions are still ongoing.