7-3-2023

Messaging Matters: Ideological Influence Online Year 3 Final Report

Matthew L. Jensen
University of Oklahoma

Shane Connelly
University of Oklahoma

Shaila Miranda
University of Oklahoma

Hairong Song
University of Oklahoma

Ares Boira Lopez
University of Oklahoma

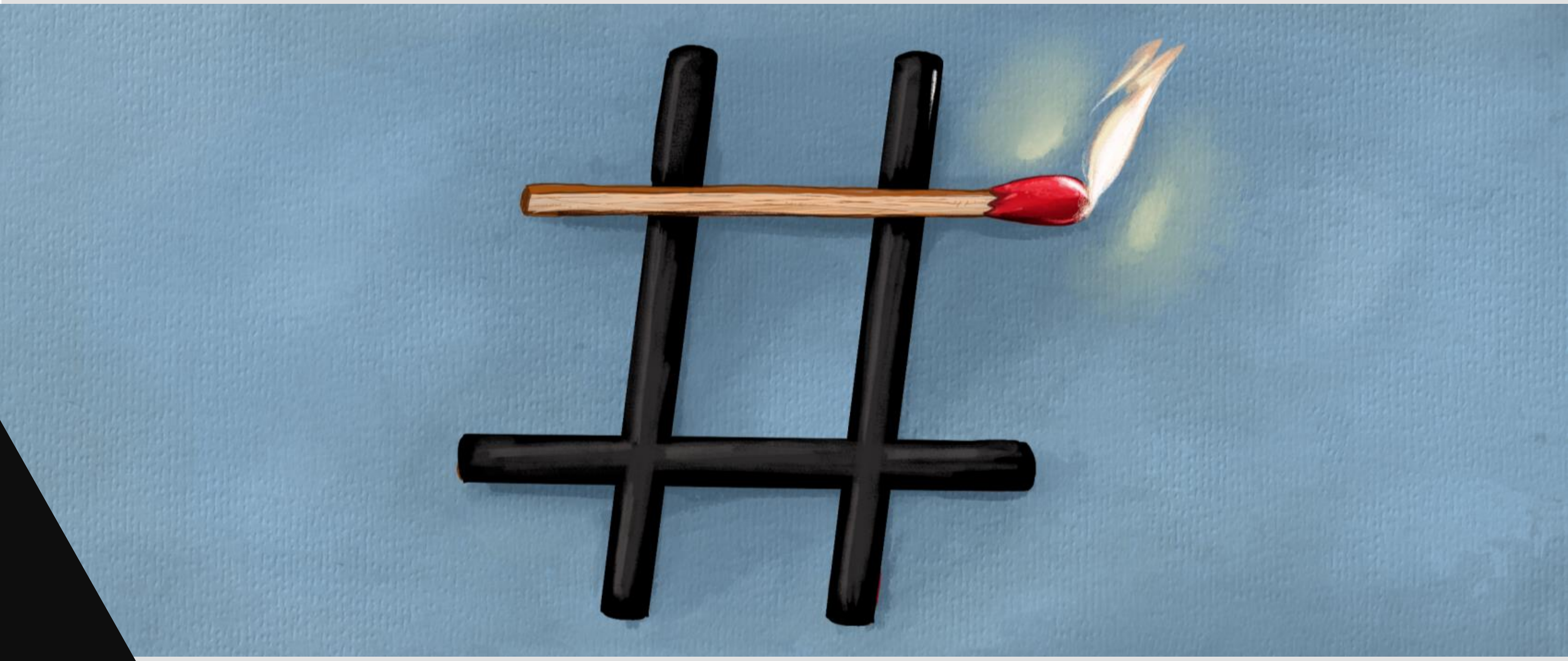
See next page for additional authors

Authors

Matthew L. Jensen; Shane Connelly; Shaila Miranda; Hairong Song; Ares Boira Lopez; Cecilia Gordon; Joseph Stewart; and National Counterterrorism Innovation, Technology, and Education Center

Messaging Matters: Ideological Influence Online Year 3 Final Report

Project Performance Reporting July 1, 2022 – June 30, 2023



July 3, 2023

Research Team

Primary Investigator
Matthew L. Jensen

Co-Primary Investigators and Consultants
Shane Connelly, Shaila Miranda, and Hairong Song

Graduate Students
Ares Boira Lopez, Cecelia Gordon, and Joseph Stewart

University of Oklahoma

NCITE
A DHS CENTER OF EXCELLENCE

About the Report

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. This year's activities 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.

Questions about this report should be directed to Matt Jensen at mjensen@ou.edu.

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 20STTPC00001-03-01. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

About the Authors

Matthew L. Jensen, Ph.D. (University of Arizona, 2007) is an Associate Professor of Management Information Systems and Associate Director of the Institute for Society and Community Transformation at the University of Oklahoma. Dr. Jensen's interests include computer-aided decision making, knowledge management, human-computer interaction, and computer-mediated communication. He studies how people attribute credibility in mediated interactions and how people filter and evaluate information they find online. Dr. Jensen has published over 100 scholarly articles in top outlets for Information Systems, Engineering, and Communication disciplines. Jensen has also been Primary Investigator or co-Primary Investigator on research projects totaling more than $9 million sponsored by federal government agencies.

Shane Connelly, Ph.D. (George Mason University, 1995) is a Professor of Industrial Organizational Psychology and Director of the Institute for Community and Society Transformation (ICAST) at the University of Oklahoma. Dr. Connelly has conducted basic and applied psychological research for over 25 years for government and non-government organizations. She has established a strong record of scholarly achievements, federal research grants, and mentorship of doctoral candidates. Her research interests include leadership and emotions, ethics, and ideological extremism in online settings. She has published over 100 articles in top peer-reviewed outlets and has received over $10 million dollars in external funding as PI or Co-PI.

Shaila M. Miranda, Ph.D., is the W.P. Wood Professor of MIS at the Price College of Business, the University of Oklahoma. She has a doctorate in Management Information Systems from the University of Georgia and an M.A. in Sociology from Columbia University. Her research focuses primarily on public discourse and shared meaning in the arenas of digital activism and innovation. She employs a combination of qualitative and computational inductive techniques. Dr. Miranda has published a book, Social Analytics, through Prospect Press, and her research has appeared in journals such as the MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Small Group Research, Information and Management, and Data Base. She serves as Senior Editor for MIS Quarterly and previously has served as Senior Editor for Information Systems Research.

Hairong Song, Ph.D. is a Professor of Quantitative Psychology at the University of Oklahoma. Her research in quantitative methods lies in two areas: (a) intensive longitudinal data analysis and dynamical system analysis, including (single- and multi-level) dynamic factor analysis and time series analysis, and (b) psychometric evaluation of behavioral measures with Frequentist and Bayesian approaches. Her current areas of substantive research focus on substance uses and (mental) health outcomes among children, adolescents, and young adults.

Ares Boira Lopez, M.A., Cecelia Gordon, M.S., and Joseph Stewart, M.S. are Ph.D. students in their fourth, third, and third years, respectively, in the Industrial-Organizational Psychology program at the University of Oklahoma. Their research interests include leadership, ethical decision-making, emotions and ideological extremism in online settings.

About NCITE

The National Counterterrorism Innovation, Technology, and Education (NCITE) Center was established in 2020 as the Department of Homeland Security Center of Excellence for counterterrorism and terrorism prevention research. Sponsored by the DHS Science & Technology Office of University Programs, NCITE is the trusted DHS academic consortium of more than 50 researchers at partner institutions across the U.S. and Europe. Headquartered at the University of Nebraska at Omaha, NCITE is a leading U.S. academic partner for counterterrorism research, technology, and workforce development.

Table of Contents

Executive Summary

Purpose
This project examines messaging strategies on publicly accessible microblogs (e.g., Twitter - prior to this platform's transition to X) used by extremist ideological groups. The primary purpose of this effort is to provide Department of Homeland Security (DHS) decision-makers and associated partners with insights about online communication and influence processes used by extreme ideological groups in social media and other publicly accessible microblog settings. By design or unknowingly, communication by these groups contains language patterns, moral disengagement (i.e., strategies for justifying/rationalizing unethical behavior), emotional appeals, and other characteristics used to recruit members, harness social identities (sense of who one is based on group membership(s) or feelings of connection), mobilize communication around issues, increase commitment to extremism, and incite violent action.

Methods
We analyze digital traces (e.g., websites, microblog archives) through automated linguistic analysis software that enables the use of built-in and customized scoring of group tweets or other online communication and messages. We also conduct coding by hand to measure certain communication characteristics that are difficult to assess using automated scoring. We examine causal effects of exposure to various communication features by designing carefully controlled, randomized studies on people's perceptions and attitudes about ideologically extreme views, emotional reactions, extent and accuracy of recall of the ideological communication exchange, and behavioral intentions with respect to the topic.

Year 3 Key Findings
Results from our Year 3 studies and analyses have uncovered the following insights:

Key Findings from Digital Trace Results

- A sample of 34 violent domestic extremist groups engaged nine types of violence, as confirmed in legitimate news reports, with physical intimidation occurring to the greatest extent and multiple deaths occurring to the least extent.
- An analysis was conducted to identify groups with similar violence profiles (i.e., cluster analysis) revealing three distinct clusters of violent domestic extremist groups 1) moderate violence, no deaths (N=4) 2) low violence of all types (N=27), 3) high physical and verbal violence, including injuries and death (N=3).
- Violent and non-violent groups were highly similar in the top 10 grievances (real or imagined wrongs or unfair treatment) word content that emerged in analysis of 24 violent and 38 non-violent groups with tweet data.
- Similar patterns also emerged in the top 10 emotions, motives, and cognitive appeals across violent and non-violent groups.
- A time series analysis of tweets of violent and non-violent groups prior to the Jan. 6, 2021 insurrection (violent event) showed similar patterns of means on grievances, emotions, cognitive appeals, motives, and moral foundations prior to Jan. 6 and in the two weeks that followed. More daily variation on these variables was observed within and across the violent and non-violent groups was observed following Jan. 21 through Feb. 25 (end of the time period examined). Interesting trends are described in this report.

Key Findings from Social Media Experiment Results

The first Year 3 experiment manipulated the presence or absence of moral disengagement, misinformation (sharing and spreading of false, inaccurate, or misleading information), and counters to moral disengagement in a social media exchange, creating eight conditions or variations in extreme ideological messaging about voter ID requirements.

- Exposure to a simulated set of social media messages about voter ID requirements resulted in participants feeling more strongly about the topic, regardless of their stance or what message variation they saw. Strength of attitude related positively to important behavioral intentions about voting and disseminating the messages.
- The presence of moral disengagement but no misinformation resulted in more trust in participant responses to the message feed and resulted in higher ratings of accuracy of the messages than when misinformation was present.
- Misinformation resulted in less liking of the pro-voter ID tweets, but only when misinformation was perceived as such.
- Participants who saw misinformation and no counter-messaging had the highest intentions to retweet pro-voter ID tweets.

The second Year 3 experiment manipulated the presence or absence of two different social identities (extreme nationalist and military) and uncertainty reduction content in a social media exchange, creating eight conditions or variations in extreme ideological messaging about voter ID requirements. All messages in this exchange were oriented toward a pro-voter ID stance.

- Exposure to a simulated set of pro-voter ID social media messages resulted in more anti-voter ID attitudes, on average, across all conditions. However, certain conditions showed increases in pro-voter ID attitudes.
- Participants who saw messages with both social identities and no uncertainty reduction content showed a shift to more pro-voter ID attitudes after seeing the messages. Those who saw both identities and uncertainty reduction content shifted to more anti-voter ID attitudes.
- Exposure to the military social identity resulted in more anger, trust, and positive affect in responding to the messages exchange, although we don’t know the specific content of what these emotions connected to.
- Exposure to the military social identity resulted in higher ratings of credibility than exposure to both social identities, which resulted in the lowest credibility ratings.
- Exposure to the military social identity and uncertainty reduction content also resulted in high credibility ratings compared to military social identity and no uncertainty reduction content, which produced modest credibility ratings that were also the lowest across all conditions.
- Exchanges containing uncertainty reduction content resulted in higher ratings of accuracy than those that did not contain this content.
- These particular social identities and uncertainty reduction content were not related to dissemination variables.

Implications

The Year 3 results have the following implications:

Digital Trace Data

- Involvement in high levels of violence was limited to a small cluster of groups. Surveillance and enforcement resources are likely best utilized focusing on this cluster.
- The language used in the lead up to violent events was very similar between violent ideological groups and non-violent ideological groups. This suggests that using rhetoric to pinpoint details about violent action yet to take place will be extremely difficult (and potentially unreliable). However, it is an open question as to whether the language similarities suggest non-violent extremist groups have the potential to turn violent.
- There are significant differences in online rhetoric between violent and non-violent groups after violent events take place. During this time, groups embracing violence are likely easier to identify. This time may also be an important target period for interventions and counter messaging.

Social Media Experiments

- Exposure to messages induces stronger attitudes about the topic, suggesting that as people encounter messaging around controversial topics, their stance on the topic will likely become more extreme (based on their pre-existing attitudes). This finding raises difficult questions for social media companies and others about the role of content moderation in discussion around highly controversial topics, especially when there is evidence of misinformation.
- Misinformation damages perceptions of accuracy and undermines sharing, but only when people notice it. If they do not detect it, people accept misinformation as legitimate evidence. This finding emphasizes the importance of identifying misinformation and helping others to identify misinformation.
- Invoking social identities in messages about controversial practices alters how people interpret and feel about the messages. This finding suggests that strategic disclosure of social identities can effectively be deployed to support or to counter extreme online rhetoric.
- Reducing uncertainty by providing verified facts in an unbiased way contributes to higher perceptions of accuracy and credibility of messages about ideological topics in social media. This technique shows promise as a tactic to counter extremist rhetoric on online platforms.

Summary of Year 3 Effort

During Year 3, we engaged in two streams of data collection and analysis. The first was the study of digital trace data collected from Twitter and mainstream news sites, 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 addressed the following questions:

Trace Data Questions

- What types of violence do extreme ideological groups engage in, and what was the incidence of each violence type for the groups of interest here?
- Are there clusters of violent ideological groups, and what are their patterns of violence?
- What grievances do violent and non-violent domestic extremist groups express the most in their Twitter messaging?
- What emotions, cognitive appeals, and motives content do violent and non-violent domestic extremist groups express the most in their Twitter messaging?
- What dissemination tactics do violent domestic extremist groups use prior to and following violent incidents in their Twitter messaging?

We also conducted two controlled, randomized experiments to examine the potential causal effects of phenomena previously observed (or anticipated) in the trace data. We addressed the following sets of questions in these studies:

Experiment Research Questions - Set 1

- How does moral disengagement in social media messages influence attitudes, emotions, recall, credibility, and dissemination intentions?
- What is the effect of moral disengagement in social media messages on these outcomes when coupled with misinformation?
- Do the effects of moral disengagement and misinformation on outcomes of interest change when a user “calls out” or counters the moral disengagement?

Experiment Research Questions - Set 2

- What are the effects of a single social identity in social media messages versus two different social identities on attitudes, emotions, recall, credibility, and dissemination intentions?
- Do the effects of single and dual social identities change when uncertainty about the topic of discussion in the social media exchange is reduced?

The sections below present details of the findings related to these research questions.

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.

Bite Back: This dataset is comprised of microblog posts from environmental activists on a platform called Bite Back. The posts described activist activities including violent and criminal acts (e.g., animal liberation, vandalism, sabotage, and arson). We did not conduct any analyses with this dataset in Year 3.

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. We did not conduct any analyses with this dataset in Year 3.

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. We did not conduct any analyses with this dataset in Year 3.

Domestic Ideological Groups (DIG): This dataset contains tweets from 63 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 American right-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 mainstream newspaper articles for mentions of violent events and group names and manually verifying the groups' participation in each event. This enabled classification of the domestic ideological groups as either violent or non-violent. Extreme ideological groups were classified as violent if they participated in at least one of the following types of violence: multiple deaths, multiple injuries, single death, single injury, property damage, physical intimidation, verbal intimidation, cyber-intimidation, or cyber-assault.

Protection of Civil Liberties in Trace Data

Archival analyses performed in Year 3 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 users mentioned in tweets. Thus, analyses were performed on anonymized social media data. Additionally, no verbatim tweets or messages were (or will be) reproduced in published manuscripts or presentations.

Trace Data Research Questions

- What types of violence do extreme ideological groups engage in, and what was the incidence of each violence type for the groups of interest here?
- Are there clusters of violent ideological groups, and what are their patterns of violence?
- What grievances do violent and non-violent domestic extremist groups express the most in their Twitter messaging?
- What emotions, cognitive appeals, and motives content do violent and non-violent domestic extremist groups express the most in their Twitter messaging?
- What dissemination tactics do violent domestic extremist groups use prior to and following violent incidents in their Twitter messaging?

Types of Violence

The violent events dataset revealed nine different types of violence associated with 34 of the extreme ideological groups from the DEG dataset, with physical intimidation occurring the most and multiple deaths occurring the least as shown in Table 1.

Table 1
Means and standard deviations for each type of violence for Violent Domestic Extremist Groups

Types of Violence	Example	Total Count	*M*	*SD*
Verbal intimidation	"pre-shooting, racial slurs were made toward black man"	50	1.47	2.63
Physical intimidation	"spraying silly string and sprinkling glitter on counter protesters"	101	2.97	4.28
Cyber-intimidation	"posting on forums about a 'pig roast' (killing of police officers)"	21	.62	1.231
Cyber-assault	"swatting conspiracy, false reported claims of violent activity"	7	.20	.69
Property damage	"bombing of a Los Angeles theatre"	35	1.03	2.14
Single injury	"assaulting a fellow protestor"	32	.94	1.54
Multiple injuries	"cuts and bruises of 14 individuals resulting from a knife fight with counter protesters"	26	.76	1.48
Single death	"Unite the Right Rally, death of Heather Heyer"	13	.38	.60
Multiple deaths	"slaying of seven Hanafi Muslims"	4	.12	.33

Note. N = 34. "Non-violent" groups were not included in calculating incident rates.

Cluster Analysis of Violent Groups

Using these violence types, we performed a cluster analysis to examine which groups looked similar in terms of their violence profiles. This analysis resulted in the identification of three clusters. The first cluster contained four groups and was characterized by a moderate level of violence, with an emphasis on physical intimidation. While deaths were not associated with the violence in this cluster, these groups have seen significant increases in violence since 2016 and were labeled "moderate ascendant." The second cluster, labeled "low and steady", contained 27 groups that were generally less violent but committed all crime types to a small degree and showed no steep increase in violence since 2016. The third cluster contained three groups and was labeled "the new high" since these groups emerged after 2016 and were characterized by higher levels of violence, including intimidation, injuries, and deaths. Figure 1 depicts mean differences across clusters and shows levels of pre-2016 and post-2016 violence. Table 2 shows the list of violent groups by cluster and Table 3 shows the violence means for these clusters.

Table 2
List of Violent Domestic Extremist Groups in Each Cluster

Cluster	Groups
Moderate ascendant	Black Bloc, Oath Keepers, Patriot Prayer, Three Percenters
Low and steady	American Freedom Party, Animal Liberation front, Aryan Nations, Atomwaffen Division, Black Hebrew Israelite, Blood and Honour, Boogaloo Bois, Bundy Ranch, Center for Security Policy, Communist Party USA, Family Research Council, Fraternal Order of Alt-Knights (FOAK), Jewish Defense League, John Brown Gun Club, Keystone United, League of the South, Moorish Soverign Citizens, Nation of Islam, National Alliance, National Socialist Movement, Redneck Revolt, Republic of New Africa, Rise Above Movement, Traditionalist Worker Party, United Constitutional Patriots, Westboro Baptist Church, Youth Liberation Front
The new high	Antifa, Proud Boys, White Lives Matter

Figure 1
Mean scores on different types of violence for the three violent group clusters.

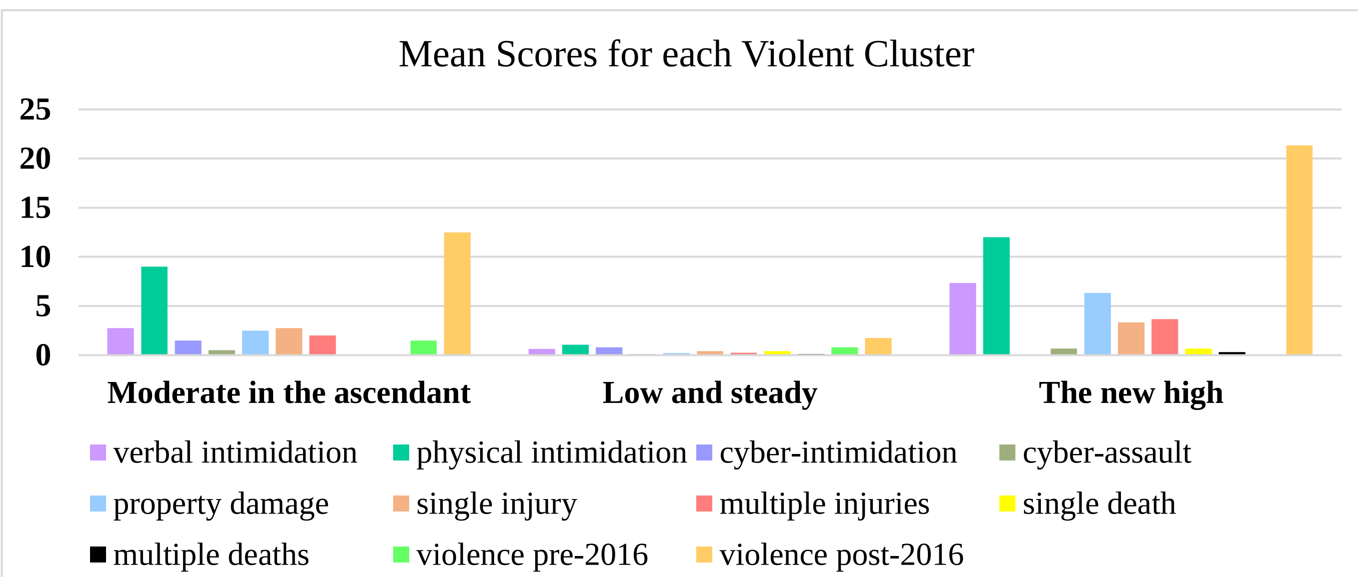


Table 3
Means and standard deviations for each type of violence for the three violent group clusters.

	Moderate in the ascendant (N=4)		Low and steady (N=27)		The new high (N=3)	
	M	*SD*	*M*	*SD*	*M*	*SD*
verbal intimidation	2.75	1.71	.63	.79	7.33	6.11
physical intimidation	9.00	4.97	1.07	1.27	12.00	1.00
cyber-intimidation	1.50	1.73	.30	.72	2.33	2.52
cyber-assault	.50	1.00	.11	.58	.67	1.15
property damage	2.50	3.00	.22	.42	6.33	2.08
single injury	2.75	1.71	.41	.69	3.33	3.06
multiple injuries	2.00	1.41	.26	.59	3.67	3.06
single death	.00	.00	.41	.64	.67	.58
multiple deaths	.00	.00	.11	.32	.33	.58
violence pre-2016	1.50	3.00	.81	1.14	.00	.00
violence post-2016	12.50	3.00	1.74	1.81	21.33	2.08

Note. N = 34.

Only 24 of these violent groups had data from Twitter, which reduced the N sizes for these clusters even further. Accordingly, for the remaining trace data analyses, we compared the 24 violent groups against the 38 non-violent group that had Twitter data.

Grievances of Violent and Non-violent Groups

We analyzed tweets from the violent and non-violent extremist groups in the DIG dataset to identify grievances these groups mentioned the most. The grievance dictionary developed by van der Vegt, Mozes, Kleinberg, and Gill (2021) is a psycholinguistic dictionary to understand grievance-fueled rhetoric for the purposes of threat assessment. The grievance dictionary can be used in the Linguistic Inquiry Word Count (LIWC) software to assess a variety of specific grievances. Definitions for the grievance constructs LIWC measured using the dictionary scoring method are shown in Appendix B. Table 4 shows the results from an analysis of the top 10 grievances for these two categories of extremist groups. Interestingly, there were no significant differences in the means for the top 10 grievances, although there were slight differences in the rank ordering across the violent and non-violent groups. Extremist groups discuss grievances related to relationships, murder, surveillance, violence, soldier, help, planning, threat, hate, and weaponry. These themes relate to armed violence, threats, and hate as well as actions related to potential responses to violence and threats such as planning and surveillance.

Table 4
Grievances of Violent and Non-violent Domestic Extremist Groups

	Violent Groups (N=24)			Non-Violent Groups (N=38)		
Top 10 Grievances	Rank order	M	SD	Rank order	M	SD
Relationship	(1)	8.31	2.83	(1)	8.25	2.17
Murder	(2)	2.65	0.69	(3)	2.41	0.85
Surveillance	(3)	2.49	0.67	(2)	3.10	2.71
Violence	(4)	2.10	0.64	(4)	2.04	0.76
Soldier	(5)	1.81	0.88	(5)	1.93	0.74
Help	(6)	1.72	0.41	(7)	1.62	0.43
Planning	(7)	1.58	0.40	(6)	1.63	0.44
Threat	(8)	1.46	0.52	(9)	1.34	0.61
Hate	(9)	1.44	0.55	(10)	1.24	0.65
Weaponry	(10)	1.43	0.54	(8)	1.60	0.72

Note. N = 62 (Violent N = 24; Non-Violent N = 38). Scores for Grievances variables represent the percentage of words identified associated with the variable. Highlights reflect where rank order differed for non-violent groups. There were no significant mean differences across violent and non-violent groups.

Emotion, Cognitive Appeals and Motives of Violent and Non-violent Domestic Extremist Groups

We also examined the top 10 emotion, cognitive appeals, and motives content expressed in the tweets of violent and non-violent extremist groups. Here again, there were no significant mean differences across the violent and non-violent groups, however, there were a few differences in rank order as shown in Table 5. Negative tone and negative emotions were the top two most prevalent themes, followed by positive tone and the motives of power and affiliation. Extreme ideological group messaging on Twitter is more negative than positive, and these groups appear to offer pathways for connecting through affiliation and social relationships and seek to gain or exert power. Several cognitive appeals also made the top 10 list, including differentiation, insight, discrepancy, tentativeness, and causation. Both types of extreme ideological groups (violent and non-violent) seek to differentiate their group from other groups, engaging in in-group and out-group kinds of behavior to do this (Connelly et al., 2015). They can accomplish this by providing insights and explanations to affiliates and by highlighting discrepancies in the arguments or ideas of out-groups. Their communications also express some degree of tentativeness or uncertainty.

Table 5
Emotions, Cognitive Appeals, and Motives of Violent and Non-violent Domestic Extremist Groups

	Violent Groups (N=24)			Non-Violent Groups (N=38)		
Top 10 Emotional, Cognitive, and Motivational Content - LIWC22						
	Rank order	M	SD	Rank order	M	SD
Negative Tone	(1)	6.76	2.10	(1)	6.62	2.15
Negative Emotions	(2)	5.00	2.42	(2)	5.10	2.11
Positive Tone	(3)	2.66	0.73	(4)	2.45	0.87
Power	(4)	2.46	0.82	(3)	2.63	0.90
Affiliation	(5)	1.74	0.72	(5)	1.65	0.69
Differentiation	(6)	1.67	0.54	(6)	1.57	0.62
Insight	(7)	1.30	0.43	(7)	1.29	0.41
Discrepancy	(8)	0.98	0.40	(9)	0.91	0.38
Tentativeness	(9)	0.97	0.41	(8)	0.93	0.46
Causation	(10)	0.92	0.26	(10)	0.89	0.31

Note. N = 62 (Violent N = 24; Non-Violent N = 38). Scores for the emotional, cognitive, and motivational appeals represent the percentage of words identified associated with the variable. Highlights reflect where rank order differed for non-violent groups. There were no significant mean differences across violent and non-violent groups.

Dissemination Before and After a Violent Event

In order to examine violent and non-violent extremist group dissemination tactics in Twitter communications prior to and following a violent incident, time series plots were created for violent and non-violent groups before and after the Jan. 6, 2021 attack on the U.S. Capitol building. Daily means for each of the grievances, emotions (LIWC and Plutchik), cognitive appeals, motivations, and moral foundations were calculated and graphed for the violent and non-violent groups from Dec. 3, 2020 through Feb. 25, 2021. Graphs showing trends of larger differences across the violent and non-violent groups are shown below. Appendix C provides an expanded set of time series graphs for key grievances, emotion, cognitive appeals, and motives.

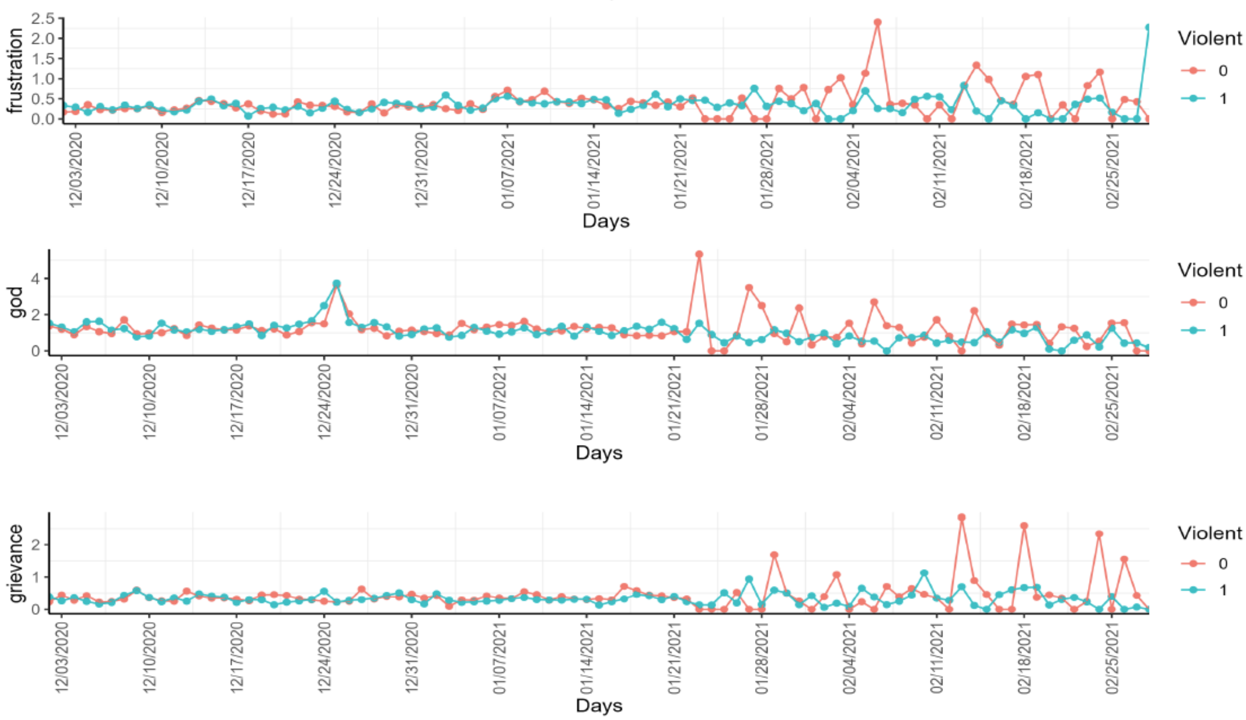
Overall, the time series analyses showed a high degree of similarity for violent and non-violent groups on all variables in the 30 days leading up to Jan. 6, 2021 and in the 14 days after Jan. 6. Interestingly, variation within and across the violent and non-violent groups increases substantially after Jan. 21 (the Presidential inauguration) and through the end of the time period sampled, Feb. 25, 2021.

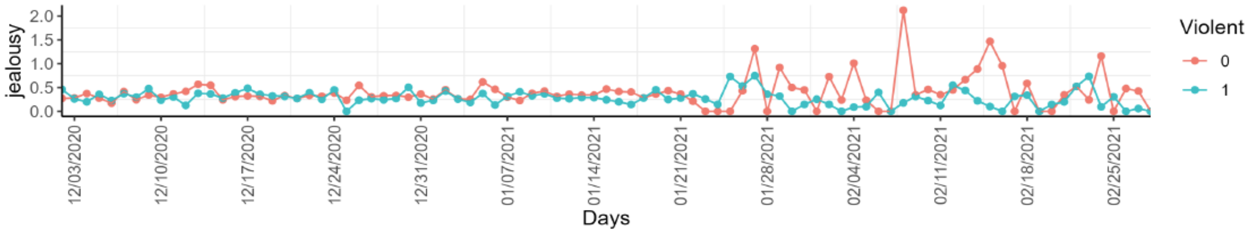
As shown below in Figure 2, non-violent groups showed more spikes in frustration, references to God (or higher power), complaints or feelings of injustice, envy or resentment of others, and planning or making detailed arrangements or strategies for a response. Interestingly, violent groups showed a slightly higher level of relationship connections or associations. Violent and non-violent groups used words related to the use of physical force or aggression to cause harm or damage, but the non-violent groups saw more daily fluctuation.

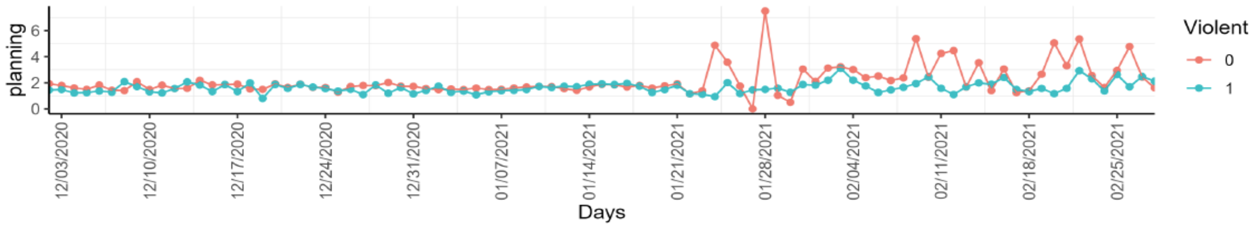
Violent groups had higher levels of negative and positive emotional tone, as well as negative emotions when assessed using the LIWC dictionaries. When overall intensity of negative and positive emotions was assessed using the Plutchik scoring approach, non-violent groups showed higher intensity levels than violent groups. Non-violent groups also had more surges of certainty, achievement, and power than violent groups.

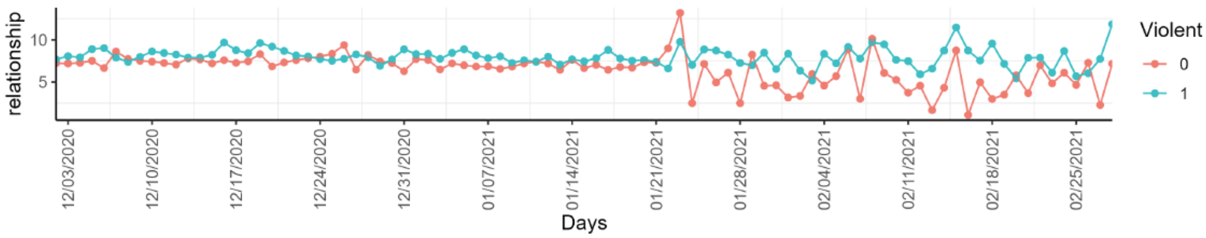
With regard to the use of words related to moral foundations in their tweets, violent groups' content indicated they saw the care foundation as more of a vice than non-violent groups, whereas non-violent groups saw the fairness and sanctity foundations as virtues more than violent groups did.

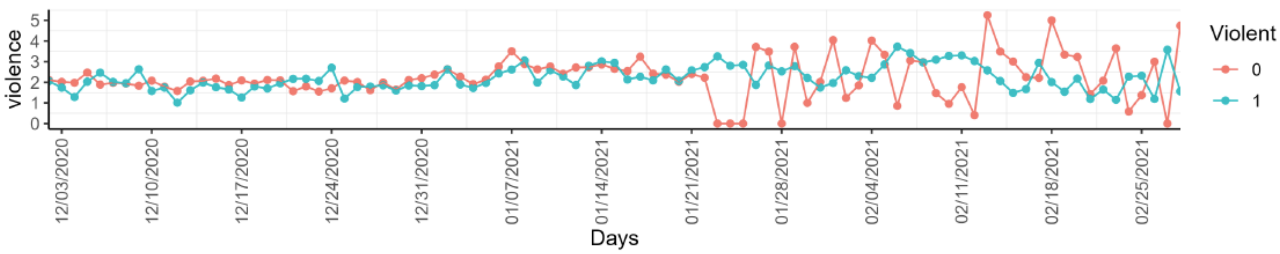
Figure 2.
Time Series Graphs Showing Different Trends for Violent and Non-violent Group Tweets Before and After the Jan. 6, 2021 Attack on the U.S. Capitol Building

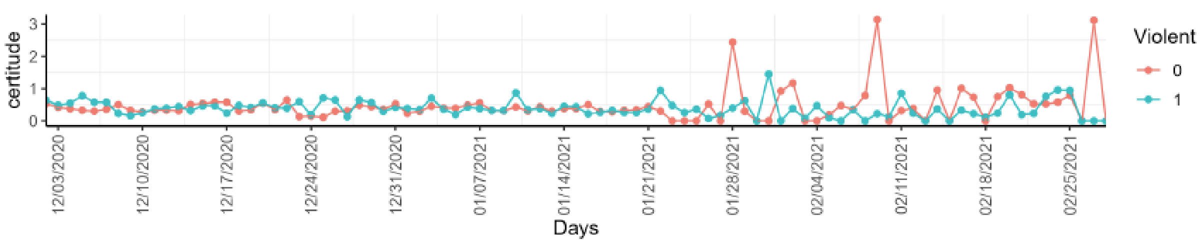


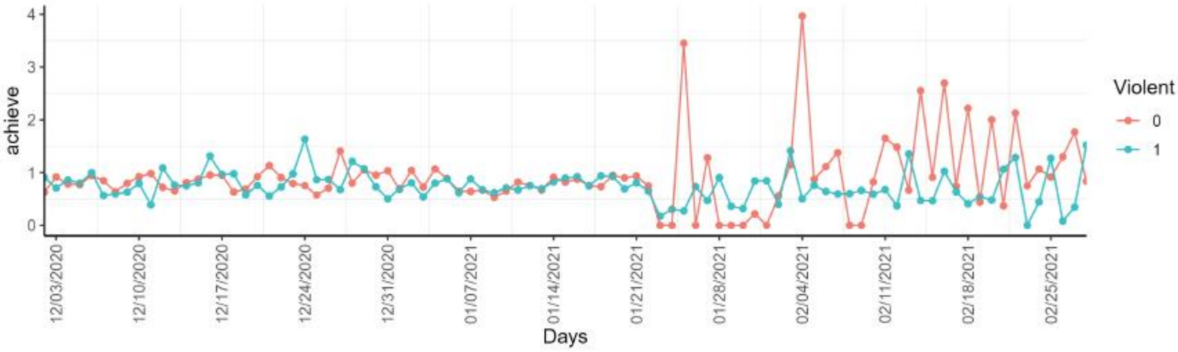

jealousy
2.0
1.5
1.0
0.5
0.0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

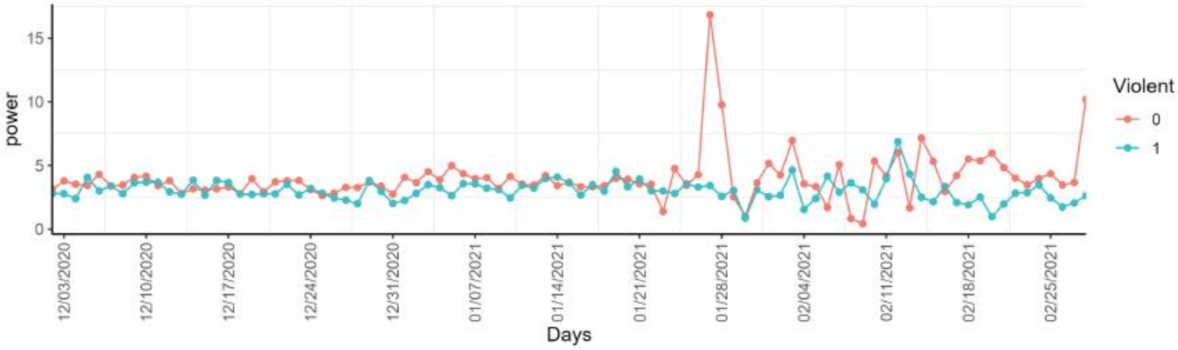

planning
6
4
2
0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

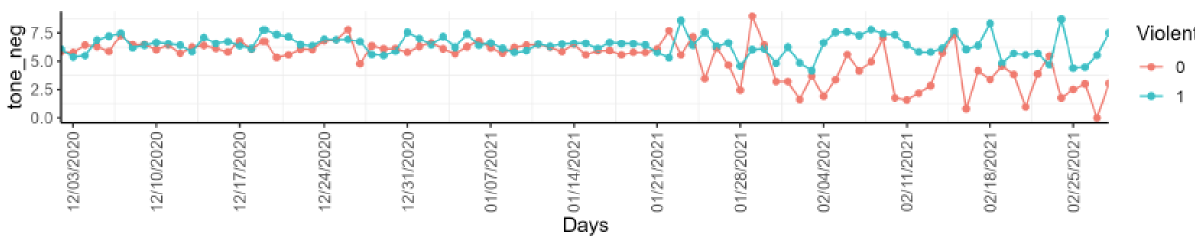

relationship
10
5
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

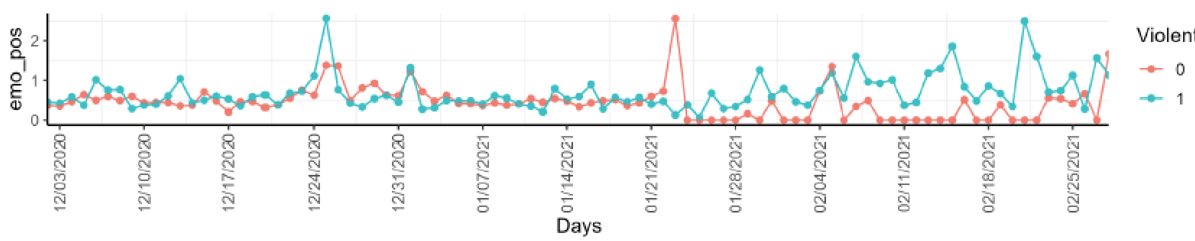

violence
5
4
3
2
1
0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

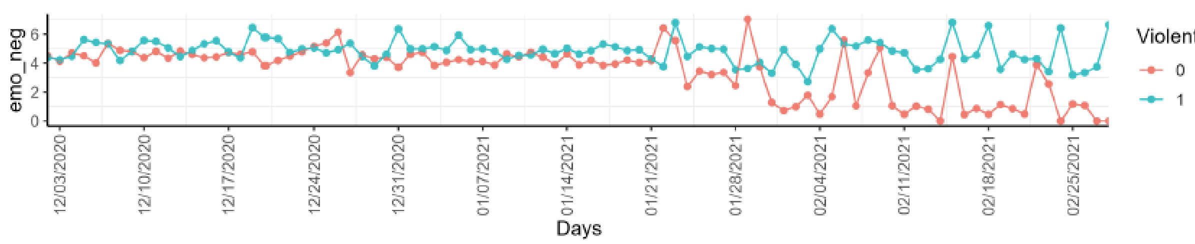

certitude
3
2
1
0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

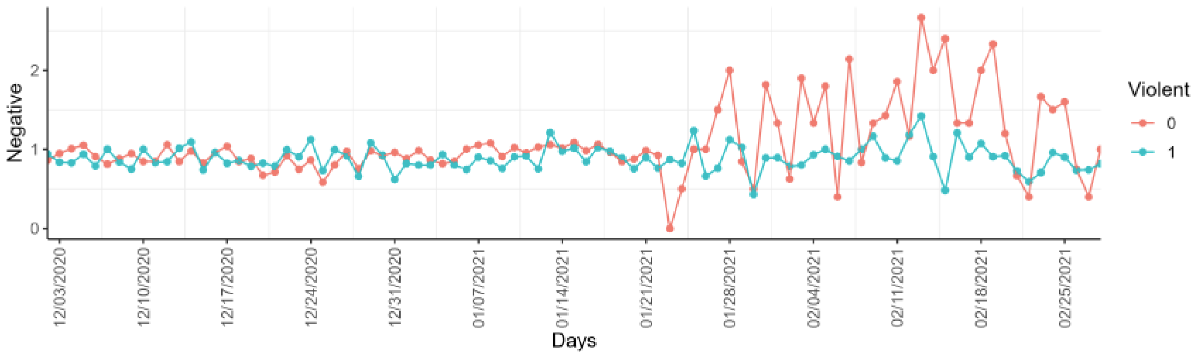

achieve
Days
Violent
0
1

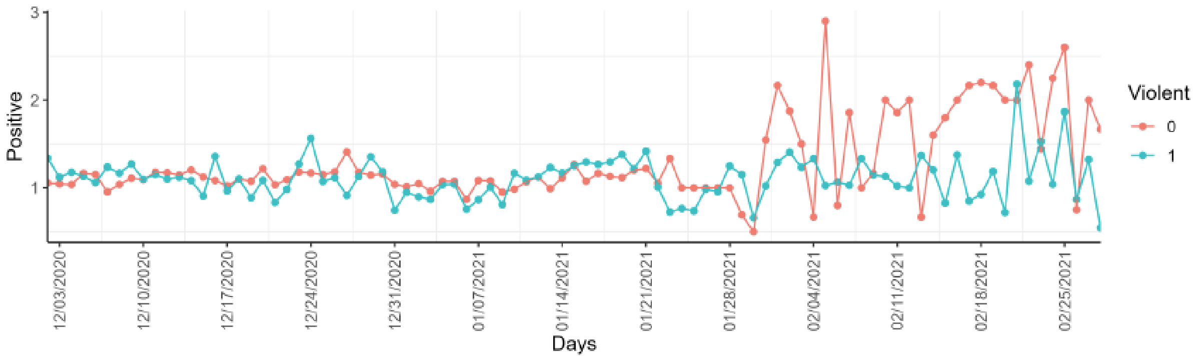

power
Days
Violent
0
1

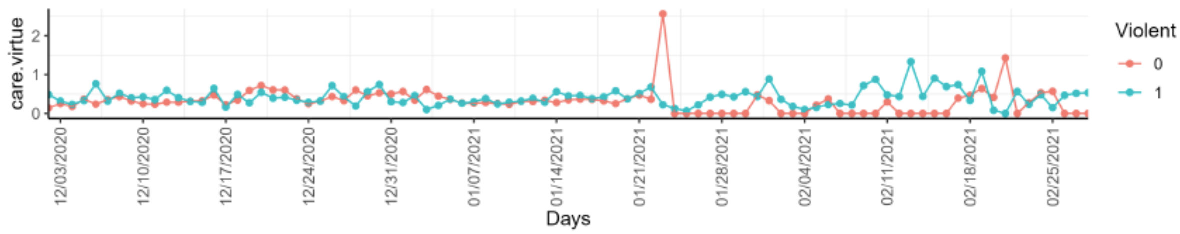

tone_neg
Days
Violent
0
1

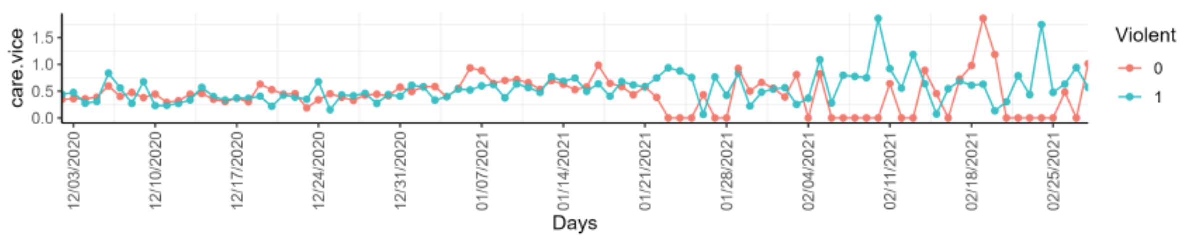

emo_pos
Days
Violent
0
1

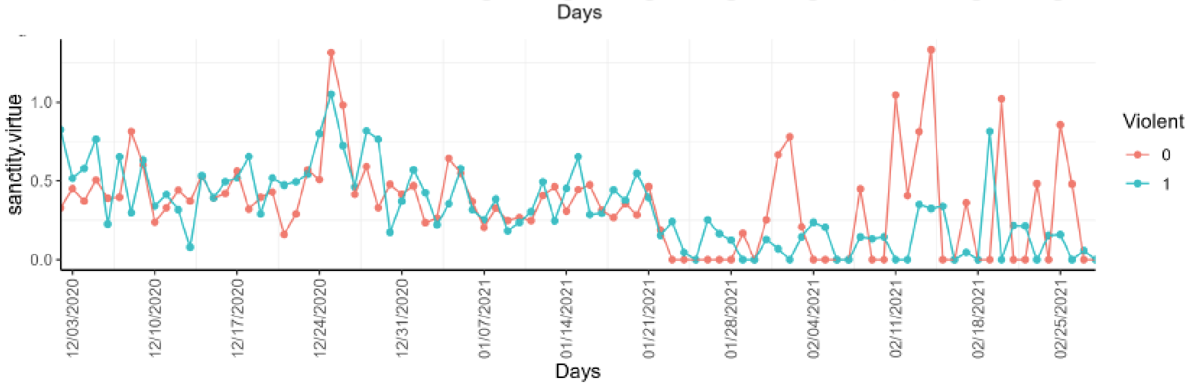

emo_neg
Days
Violent
0
1


Negative
0
1
2
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1


Positive
1
2
3
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1


care.virtue
0
1
2
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1


care.vice
0.0
0.5
1.0
1.5
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1


sanctity.virtue
0.0
0.5
1.0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

Implications for Practice

Involvement in high levels of violence (damage to property, physical harm, loss of life) was relatively limited to a small cluster of groups. Surveillance and enforcement resources are likely best utilized focusing on this cluster. However, this cluster recently formed and is growing.

The language used in the lead up to violent events was very similar between violent ideological groups and non-violent ideological groups. This suggests that using rhetoric to pinpoint details about violent action yet to take place will be extremely difficult (and potentially unreliable). Future analysis will also likely be complicated by shifting platform policies (e.g., allowable speech), membership demographics, and posting behavior, rendering models for analyzing rhetoric based on past language less applicable to current rhetoric.

There are significant differences in online rhetoric between violent and non-violent groups after violent events take place. This finding suggests that the time following violent events is especially noteworthy as groups attempt to explain their position in relation to (and potentially in defense of) violence that took place. Groups embracing violence are likely easier to identify. This post-event time may also be an important target period for interventions and counter messaging.

Experimental Studies and Findings

Voter ID Study II – Moral Disengagement, Misinformation and Counter-Messaging

This study posed the following questions:

- How does moral disengagement in social media messages influence attitudes, emotions, recall, credibility, and dissemination intentions?
- What is the effect of moral disengagement in social media messages on these outcomes when coupled with misinformation?
- Do the effects of moral disengagement and misinformation on outcomes of interest change when a user "calls out" or counters the moral disengagement?

This randomized controlled online experiment examined how the presence of moral disengagement mechanisms (i.e., rationalizing unethical behavior to make it seem acceptable) in an ideological Twitter feed about voter identification (voter ID) requirements influenced attitudes toward the topic, emotional reactions, recall of specific tweets, credibility of the tweets, and dissemination intentions (like, retweet, hashtag, share). It additionally explored the enhancing effects of coupling moral disengagement with misinformation and the mitigating effects of counter-messaging. A simulated Twitter feed was created that included 12 tweets discussing voter ID perspectives. Moral disengagement, misinformation, and counter-moral disengagement messaging were manipulated across eight unique study conditions which are shown in Table 6 below.

A battery of covariate (control) measures included digital activism, political orientation, general social media usage, social desirability, and the big-5 personality variables. Coefficient alpha scales' reliability ranged from .70 to .91 for the covariate and outcome measures. T-tests and Cohen's d's were evaluated to assess differences in importance of the issue, stance, and attitude strength before and after exposure to the tweets. A series of Analysis of Covariance (ANCOVA) analyses were conducted to examine main and interactive effects of moral disengagement, misinformation, and counter-messaging 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. Correlation analyses were conducted to evaluate the relationship of attitude variables to dissemination behaviors. See Table 7 for full correlation results.

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 eight conditions, exposure to this social media exchange resulted in increases in participants feeling more strongly about the topic ($t(247) = -2.09$, $p < .05$, $d = .82$). In turn, changes in strength of attitude toward voter ID requirements was positively related to credibility perceptions of the feed exchange ($r=.171$, $p < .01$) and feelings of encouragement to vote, gather information about voting, and discuss voting with others ($r = .160$, $p < .01$). Changes in strength of attitude was also positively related to intentions to retweet ($r = .148$, $p < .01$), hashtag ($r = .164$, $p < .01$), and share the content with similar others ($r = .221$, $p < .01$) and dissimilar others ($r=.123$, $p<.05$).

Exposure to the social media exchange also moved participants towards an anti-voter ID stance ($t(247) = -2.17$, $p < .05$, $d = .97$). Counter-messaging was positively associated with changes in stance toward a more anti-voter ID attitude ($F(1,236) = 9.20$, $p < .01$, $\eta p2 = .04$).

Table 6.
Manipulations, Conditions, and Example Tweets – Moral Disengagement, Misinformation, and Countering

	Manipulations		
Condition	Moral Disengagement	Misinformation	Moral Disengagement Counter
1			
2			
3			
4			
5			
6			
7			
8			
Example Tweets	<u>Present:</u> "I don't understand this sob story about providing a photo ID to vote. In e some parts of the world certain groups can't vote. So having to show ID is not a big deal and should be the only way we vote." <u>Absent:</u> "Voter ID should be a must & only real ID should be accepted. People who claim to have trouble getting one aren't trying."	<u>Present:</u> "All states should require photo ID both to vote in person and to vote by absentee ballot. No excuses. Getting a photo ID doesn't cost anything." <u>Absent:</u> "I don't understand this sob story about providing a photo ID to vote. Having to show ID is not a big deal and should be the only way to vote."	<u>Present:</u> "Claims of voter fraud are questionable and requiring ID definitely prevents poll access for legitimate voters." <u>Absent:</u> "Questions about voting? Please contact your local election board for accurate, unbiased, nonpartisan answers."

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 moral disengagement and misinformation was associated with trust in the tweet response ($F(1, 108) = 4.47$, $p < .05$, $\eta p2 = .04$). Trust was highest when participants were exposed to messages infused with moral disengagement and no misinformation, and it significantly decreased when misinformation was present. This suggests that moral disengagement aimed at justifying an ideological stance could be a more powerful influence on trust than direct use of misinformation.

Trust in turn was positively related to various dissemination variables, including sharing with dissimilar others ($r = .124$, $p < .05$), liking the pro-voter ID tweets ($r = .158$, $p < .05$), retweeting the pro-voter ID tweets ($r = .156$, $p < .05$).

Recall & Mentioning
Participants were asked to restate as many of the tweets as they could remember. A team of three trained coders will rate on three-point scale the extent to which information from each tweet in the feed was reflected in these restatements. These will be aggregated to create recall scores. Participants' response to the tweet feed will also be evaluated for the extent to which it mentioned voter ID tweets. A team of coders is currently coding the data for recall of each tweet and for mention of specific tweets in participants' open-ended tweet response.

Credibility
Participants rated message credibility on a seven-point Likert scale assessing the trustworthiness, fairness, expertise, goodwill, and currency of the tweet content. They also rated perceptions of message accuracy on a five-point Likert scale.

The interaction between moral disengagement and misinformation was associated with perceptions of accuracy ($F(1, 238) = 4.02$, $p < .05$, $\eta p2 = .02$). Perceptions of accuracy were highest when participants were exposed to messages infused with moral disengagement and no misinformation, and lowest when misinformation was present.

Accuracy was in turn positively related to liking ($r = .180$, $p < .01$) and retweeting ($r = .144$, $p < .05$) the pro-voter ID tweets, and sharing with dissimilar others ($r = .131$, $p < .05$.)

Dissemination Intentions
Participants were asked to rate the extent to which they would like (Y/N), hashtag words (Y/N), and retweet (Y/N) each tweet. They were also asked 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).

Misinformation was negatively associated with liking the pro-voter ID tweets ($F(1, 118) = 5.29$, $p < .05$, $\eta p2 = .04$). However, it appears that that is the case when participants identity the content as misinformation, because those who perceived the tweets as more accurate had higher intentions to like the pro-voter ID tweets ($F(1, 118) = 6.10$, $p < .05$, $\eta p2 = .02$).

Participants exposed to misinformation and no counter-messaging had the highest intentions to retweet the pro-voter ID tweets, but those exposed to misinformation with counter-messaging had the lowest intentions to retweet them ($F(1, 119) = 3.64$, $p = .059$, $\eta p2 = .03$).

Table 7.

Correlations for Voter ID Study II

Correlations between Attitudes, Emotions, Credibility Perceptions, and Dissemination Intentions

	1	2	3	4	5	6	7	8	9	10	11	12
Attitudes												
1. Strength Change	1											
2. Stance Change	-.169**	1										
3. Opinion Change	.104	.244**	1									
Emotions												
4. Anticipation	.054	.096	.112	1								
5. Fear	.002	-.021	-.027	.229**	1							
6. Joy	.013	.074	.032	.775**	.246**	1						
7. Trust	.057	.000	.078	.632**	.188**	.737**	1					
8. Negative Affect	.022	.034	-.042	.575**	.596**	.501**	.422**	1				
9. Positive Affect	.065	.047	.091	.644**	.264**	.734**	.739**	.459**	1			
Credibility Perceptions												
10. Twitter Cred.	.032	.011	.128*	.061	-.082	.024	.097	.004	.048	1		
11. Feed Cred	.171**	-.151*	.026	.013	-.063	.027	.068	.022	.071	.412**	1	
12. Accuracy Perception	.089	-.198**	.010	-.025	-.002	.025	.083	.029	.038	.399**	.626**	1
13. Encouragement	.160**	.116	.215**	.114	.139*	.086	.070	.098	.155*	.198**	.144*	.094
Dissemination Intentions												
14. Like	.105	.025	.162*	.003	-.012	-.016	.015	.029	-.001	.116	.133*	.097
15. Pro-Like	.022	-.177**	-.050	.067	.028	.114	.158*	.092	.121	.018	.148*	.180**
16. Counter-Like	.094	.178**	.229**	-.064	-.036	-.121	-.125*	-.049	-.114	.119	.031	-.043
17. Retweet	.148*	-.017	.168**	.027	.039	.079	.037	.039	.055	.131*	.091	.058
18. Pro-Retweet	.081	-.185**	-.007	.082	-.017	.029	.156*	.054	.122	.106	.157*	.144*
19. Counter-Retweet	.114	.153*	.223**	-.042	.066	-.073	-.099	-.001	-.044	.070	-.030	-.061
20. Hashtag	.164**	.011	.102	.168**	.037	.157*	.096	.067	.126*	-.022	-.037	-.081
21. Share w/ Similar	.221**	-.105	.093	.059	-.019	.083	.066	.065	.123*	.233**	.222**	.053
22. Share w/ Dissimilar	.123*	-.074	.118	.142*	.038	.098	.131*	.124*	.125*	.235**	.100**	.131*
23. General Dissemination	.069	.001	.023	-.067	-.156*	-.113	-.110	-.129*	.215**	.134*	.139**	.056

Note. N=248, *Significance <.05, **Significance < .01

Table 7 continued.

Correlations between Attitudes, Emotions, Credibility Perceptions, and Dissemination Intentions

	13	14	15	16	17	18	19	20	21	22	23
Attitudes											
1. Strength Change											
2. Stance Change											
3. Opinion Change											
Emotions											
4. Anticipation											
5. Fear											
6. Joy											
7. Trust											
8. Negative Affect											
9. Positive Affect											
Credibility Perceptions											
10. Twitter Cred.											
11. Feed Cred											
12. Accuracy Perception											
13. Encouragement	1										
Dissemination Intentions											
14. Like	.171**	1									
15. Pro-Like	.024	.535**	1								
16. Counter-Like	.173**	.700**	-.223**	1							
17. Retweet	.146*	.394**	.232**	.264**	1						
18. Pro-Retweet	.002	.188**	.423**	-.143*	.638**	1					
19. Counter-Retweet	.187**	.331**	-.100	.475**	.688**	-.120	1				
20. Hashtag	.168**	.147*	.078	.092	.168**	.079	.143*	1			
21. Share w/ Similar	.247**	.109	.072	.060	.207**	.114	.160*	.210*	1		
22. Share w/ Dissimilar	.185**	.176**	.063	.149*	.241**	.108	.209**	.090	.551**	1	
23. General Dissemination	.081	-.020	.001	-.031	.029	.022	.016	.022	.109	.054	1

Note. N=248, *Significance <.05, **Significance < .01

Voter ID Study III – Nationalist Social Identity, Uncertainty Reduction and Military Social Identity

This study posed the following questions:

- How does introducing a second social identity affect the influence of an ideologically extreme social identity on attitudes, emotions, recall, credibility, and dissemination intentions?
- How about when we reduce uncertainty about the topic of discussion?

This randomized controlled online experiment examined how the presence of social identity in an ideological Twitter feed about voter ID requirements influenced attitudes toward the topic, emotional reactions, recall of specific tweets, credibility of the tweets, and dissemination intentions (like, retweet, hashtag, share). It additionally explored the mitigating effects of introducing a second social identity as well as reducing uncertainty about the topic of discussion. In our prior work, we found that tweets that included two types of social identity information (references to minority identities and political preferences) were less impactful than those containing only minority identity references. However, the prior study did not control for stance on the ideological issue, such that minority social identity tweets tended to be anti-voter ID while political social identity tweets contained both conservative and liberal perspectives, which were pro- and anti-voter ID, respectively. In the present study, the second social identity (i.e., U.S. military veteran identity) was able to be presented in an ideologically consistent way with the pro-voter ID stance. It was also developed to be less extreme than the first social identity (i.e., U.S. nationalist identity).

A simulated Twitter feed was created that included nine tweets discussing voter ID perspectives. Nationalist social identity, uncertainty reduction and military social identity were manipulated across eight unique study conditions which are shown in Table 8 below.

A battery of covariate measures included digital activism, political orientation, general social media usage, social desirability, and the big-5 personality variables. Coefficient alpha scales' reliability ranged from .72 to .94. A series of Analysis of Covariance analyses were conducted to examine main and interactive effects of social identity(ies) and uncertainty reduction 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. See Table 9 for correlation results.

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 eight conditions, exposure to the social media exchange resulted in participants moving their stance toward a more anti-voter ID requirement stance ($t(227)=-1.94$, $p=.054$, $d=1.12$). ANCOVA results revealed a significant three-way interaction between exposure to the nationalist social identity, military social identity, and uncertainty reduction tweets and certain outcomes. Those exposed to the two social identities, with no uncertainty reduction moved their stance to be more pro-voter ID requirements ($M=-.254$, $SD=.961$) whereas those exposed to the two social identities as well as the uncertainty reduction moved their stance to be more anti-voter ID requirements ($M=.337$, $SD=1.04$; $F(1,216)=6.23$, $p=.013$, $h^2=.028$). In turn, changing their feelings toward the issue of voter ID was negatively related to intentions to share voter ID content with dissimilar others ($r = -.140$, $p < .05$).

Table 8.
Manipulations, Study Conditions, and Example Tweets

	Manipulations		
Condition	Extreme Nationalist ID	Uncertainty Reduction	Service in Military ID
1			
2			
3			
4			
5			
6			
7			
8			
Example Tweets	Present: "All states should require photo ID both to vote in person and to vote by absentee ballot. It's un-American to not have ID. Many states offer one free of charge." Absent: "Voter ID laws were set up by elected officials to stop voter fraud not to block voters. More insane propaganda!"	Present: "21 million Americans do not have a photo ID. It's not just the ID cost. Document fees, travel expenses, and time are not always accessible to all." Absent: "Interesting how people always argue about this. There is definitely room for improvement on this issue."	Present: "I agree that our elections should be protected, but I did not serve in the military to see people considering force against other Americans." Absent: "Hmm...I'm still thinking about this one. I think this is an important topic to discuss."

Among those participants in conditions four and seven, the strength of their attitudes toward the issue of voter ID requirements changed. In condition four, participants were exposed to the nationalist social identity alone, and this resulted in a decrease in the strength of their stance ($t(33)=1.99$, $p=.055$, $d=.55$). In condition seven, participants were exposed to the military social identity alone, and this resulted in an increase in the strength of their stance ($t(32)=-1.79$, $p=.042$, $d=.62$). In turn, changes in one's feelings toward voter ID requirements was positively related to perceptions of Twitter credibility ($r = .128$, $p = .05$) and perceptions of exchange credibility ($r = .171$, $p < .01$). These changes were also negatively related to hash tagging the content in the exchange ($r = -.127$, $p = .05$) and positively related to encouragement to vote, gather information about voting, and discuss voting rights with others ($r = .180$, $p < .01$).

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 emotions of submission, anger, trust, and positive affect were retained for analyses, as these were the only emotions that passed the Levene's test of homogeneity of variance.

There was a main effect of exposure to the military social identity on anger ($F(1,257)=5.17$, $p=.024$, $h^2=.020$), trust ($F(1,257)=5.85$, $p=.016$, $h^2=.017$), and positive affect ($F(1,257)=5.55$, $p=.019$, $h^2=.0121$) in participant response to the Twitter feed, such that those exposed to the military social identity expressed more anger, trust, and positive affect in their responses. Additionally, there was a two-way interaction between exposure to the nationalist social identity and uncertainty reduction feed content on submission in participant responses *(F(1,252)=4.25, p=.040, h^2=.017)* such that those exposed to the uncertainty reduction content expressed the greatest amount of submission, whereas those exposed to neither the uncertainty reduction content nor the nationalist social identity expressed the least amount of submission.

Anger was negatively related to perceptions of credibility of Twitter *(r = -.130, p < .05)* and positively related to intentions to like *(r = .198, p < .01)*. Trust was positively related to intentions to like *(r = .230, p < .01)*, and intentions to engage in any type of dissemination tactic (i.e. like, retweet, etc.), *(r = .139, p < .05)*. Positive affect was positively related to feelings of encouragement to vote, gather information about voting and discuss voting rights with others *(r = .156, p < .05)*, intentions to like *(r = .213, p < .01)*, and intentions to engage in any type of dissemination tactic *(r = .133, p < .05)*. Submission was positively related to intentions to share voter ID content with similar others *(r = .148, p < .05)*.

Recall & Mentioning

Participants were asked to restate as many of the tweets as they could remember. A team of three trained coders will rate on three-point scale the extent to which information from each tweet in the feed was reflected in these restatements. These will be aggregated to create recall scores. Participants' response to the tweet feed will also be evaluated for the extent to which it mentioned voter ID tweets. A team of coders is currently coding the data for recall of each tweet and for mention of specific tweets in participants' open-ended tweet response.

Accuracy and Credibility

Participants rated message credibility on a seven-point Likert scale assessing the trustworthiness, fairness, expertise, goodwill, and currency of the tweet content. Perception of credibility of Twitter as a platform was also assessed.

There was a two-way interaction between exposure to the nationalist social identity and exposure to the military social identity on perceptions of the Twitter feed credibility *(F(1,253)=4.48, p=.035, h^2=.017)*. Those exposed to the military social identity alone perceived the Twitter feed as the most credible *(M=3.70, SD=1.02)* and those exposed to both social identities perceived the Twitter feed to be the least credible *(M=3.49, SD=1.07)*. There was also a two-way interaction between exposure to the military social identity and uncertainty reduction content on perceptions of the Twitter feed credibility *(F(1,253)=4.97, p=.027, h^2=.019)*. Those exposed to both the military identity and the uncertainty reduction content perceived the feed as the most credible *(M=3.85, SD=1.03)* and those exposed to the military social identity alone perceived the feed as the least credible *(M=3.32, SD=1.01)*. There was also a main effect of exposure to uncertainty reduction tweets on perceptions of accuracy of the feed (F(1,256)=4.48, p=.035, h^2=.017). Participants exposed to the uncertainty reduction tweets perceived the feed as more accurate (M=2.83, SD=.643) compared to those who were not exposed to the uncertainty reduction tweets (M=2.65, SD=.733).

Perceptions of feed credibility was positively related to intentions to a) retweet *(r = .128, p < .05)*, b) disseminate content with similar others *(r = .192, p < .05)*, c) disseminate content with dissimilar others *(r = .209, p < .01)*, and d) disseminate in general *(r = .215, p < .01)*. Feed credibility perceptions were also positively related to feelings of encouragement to vote, intentions to gather information about voting and intentions to discuss voting rights with others (r = .141, p < .05). Perception of credibility of Twitter in general was positively related to intentions to disseminate content with dissimilar others *(r = .194, p < .01)*. Perception of feed accuracy was positively related to intentions to retweet *(r = .128, p < .05)*, intentions to disseminate with similar others *(r = .182, p< .01)*, intentions to disseminate with dissimilar others ($r = .207$, $p < .01$), intentions to disseminate in general ($r = .222$, $p < .01$), and feelings of encouragement to vote, gather information about voting, and discuss voting rights with others ($r = .163$, $p < .05$).

Dissemination Intentions

Participants were asked to rate the extent to which they would like (Y/N), hashtag words (Y/N), and retweet (Y/N) each tweet. They were also asked 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). A general intention to disseminate variable was created using an aggregate of the standardized scores for the aforementioned variables.

There were no main or interactive effects of the social identity and uncertainty reduction conditions on any of the dissemination variables.

Table 9.
Correlations for Voter ID Study III

Correlations between Attitudes, Emotions, Credibility Perceptions, and Dissemination Intentions

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Attitudes																	
1. Strength Change	1																
2. Stance Change	.295**	1															
3. Opinion Change	-.063	-.095	1														
Emotions																	
4. Anger	.119*	.011	-.030	1													
5. Trust	.154*	-.006	.055	.550**	1												
6. Positive Affect	.151*	.047	.034	.586**	.795**	1											
7. Submission	.117	.017	-.065	.499**	.501**	.573**	1										
Credibility Perceptions																	
8. Twitter Cred.	.128*	.042	.161**	-.130*	-.068	-.063	-.017	1									
9. Feed Cred	.171**	-.055	.121*	-.061	-.001	.030	-.069	.466**	1								
10. Accuracy Perception	.142*	.007	.078	-.092	.003	-.013	-.028	.459**	.695**	1							
11. Encourage	.180**	-.026	.191**	.063	.117	.156*	.066	.107	.141*	.163**	1						
Dissemination Intentions																	
12. Like	.059	-.032	.100	.198**	.230**	.213**	.081	-.064	.063	.094	.136*	1					
13. Retweet	.057	.016	.241**	.031	.063	.079	-.038	.077	.136*	.128*	.185**	.499**	1				
14. Hashtag	-.127*	.009	-.018	-.069	-.019	.009	.013	-.030	-.005	.005	-.067	-.193**	-.142*	1			
15. Share w/ Similar	.093	-.014	.214**	.060	.087	.074	.148*	.113	.192*	.182**	.291**	.228**	.298**	-.148*	1		
16. Share w/ Dissimilar	.096	-.140*	.191**	-.006	.039	.003	-.056	.194**	.209**	.207**	.179**	.150*	.150**	-.130*	.611**	1	
17.General Dissemination	.064	-.059	.262**	.074	.139*	.133*	.041	.105	.215**	.222**	.261**	.609**	.660**	.139*	.781**	.647**	1

Note. N=263, *Significance <.05, **Significance < .01

Implications for Practice

Exposure to messages (with stances pro and against controversial practices) induces stronger attitudes about the topic. This finding was replicated across both experiments, demonstrating strong empirical support. This suggests that as people encounter messaging around controversial ideologically based topics, their stance on the topic will likely become more extreme (based on their pre-existing attitudes). This finding raises difficult questions for social media companies and others about the role of content moderation in discussion around highly controversial topics, especially when there is evidence of misinformation.

Misinformation damages perceptions of accuracy, but only when people notice it. If they do not detect it, people accept misinformation as legitimate evidence. This finding emphasizes the importance of identifying misinformation and helping others to identify misinformation. When misinformation is identified, it is also less likely to be shared.

Invoking social identities in messages about controversial practices alters how people interpret and feel about the messages. This finding suggests that strategic disclosure of social identities can effectively be deployed in support or to counter extreme online rhetoric.

Reducing uncertainty by providing verified facts in an unbiased way contributes to higher perceptions of accuracy and credibility. This technique shows promise as a tactic to counter extremist rhetoric on online platforms.

Trajectory of Findings

The following discussion synthesizes our findings by concept across the three years of this project.

Moral Disengagement and Moral Foundations

In Year 1, our initial investigation revealed that Bite Back, an animal activist platform containing violent messages, incorporated moral disengagement within their online posts. Building on this foundation, in Year 3, we orchestrated an experiment to delve into the repercussions of encountering ideological messages containing moral disengagement. The outcomes were illuminating – the presence of moral disengagement bolstered trust in consumer responses and fostered a sense of accuracy in information that was, in reality, inaccurate. A noteworthy inquiry emerged: What transpires when individuals confront moral disengagement and respond to it? For instance, when an extremist group resorts to dehumanization, a user might retort, "These individuals you label as 'rats' are human beings.'" Our Year 3 results suggest that exposure to such counter-messaging decreased intentions to retweet the content among users.

Regarding moral foundations, Year 2 exposed a divergence: Non-violent group affiliates referenced fairness virtues (e.g., equality, justice, rights), while violent group affiliates leaned toward sanctity virtues (e.g., purity, sacred, wholesome). Additionally, violent group leaders employed more authority vice keywords (e.g., subversion, disobey, disrespect). In Year 3, we focused on the references to moral foundation keywords around the Jan. 6, 2021 Capitol riot. Our findings unveiled non-violent groups emphasizing care virtue keywords (e.g., kindness, compassion, empathy) around Inauguration Day. References to sanctity virtue keywords surged for both violent and non-violent groups before and after the Jan. 6, event, particularly persistent for the latter. The Year 3 time series graphs illuminated the evolving references to these moral foundations, indicating shifts contingent on events.

Issues and Social identity

In Year 1, our exploration revealed that ideological groups frequently intertwined their issues with social identity cues like "we" and "they" in their social media feeds, an approach that, supported by emotional and cognitive appeals, would amplify the persuasive impact of their messages. The subsequent categorization of these groups into "violent" and "non-violent" in Year 2 allowed for a more in-depth examination. This analysis provided preliminary evidence suggesting that violent groups primarily employed negative emotional appeals, as well as insight and differentiation, when referencing social identity. Conversely, non-violent groups appeared to focus primarily on positive emotions (excluding fear) and utilized all forms of cognitive appeals when referencing social identity.

Upon closer examination, it became apparent that several key groups, particularly violent ones, had significant gaps in their data due to affiliated Twitter accounts being suspended. Consequently, in Year 3, we conducted additional research to locate updated Twitter handles for these users and gather more data. When we reran the same analysis with this additional data, most of the differences between violent and non-violent groups disappeared. Both types of groups employed positive and negative emotions, as well as cognitive appeals, when referencing social identity. The primary distinguishing factor was their use of certainty. While violent groups leaned more heavily on certainty in their general communications, non-violent groups utilized it significantly more when referencing social identity. Another noteworthy finding was that violent ideological groups referenced social identity cues to a greater extent than their non-violent counterparts.

To examine the effects of linking social identity with an issue, we undertook two experimental studies in Year 2. One study centered on the death penalty, intertwining the social identities "minority" and "pro-life," while the other involved voter ID requirements and social identities "minority" and "political." Preliminary evidence indicated that invoking inherent identities (e.g., "minority") evoked stronger negative emotions compared to chosen social identities (e.g., "political" and "pro-life") or a combination of the two. Notably, inherent identity references led to increased information processing depth and dissemination intentions. However, introducing chosen social identities alongside inherent ones mitigated these effects, diminishing content recall and sharing desire.

Building on these insights, Year 3 extended this research, exploring strategies to diminish the persuasive impact of ideological messaging. We investigated adding a second social identity and a non-ideological tweet response providing factual information to reduce uncertainty. The focal issue in the experiment was voter ID requirements, with all the tweets adopting a pro-voter ID stance, and the identities were “extreme nationalist” and “military.” Although exposure to both social identities influenced people to adopt a more pro-voter ID position, uncertainty-reducing counter-messages were observed to moderate these effects, nudging users toward anti-voter ID positions.

The effects observed in our previous Year 2 studies could not be replicated in our Year 3 study due to the absence of an inherent social identity. Nevertheless, there is preliminary evidence indicating that the impact of social identity varies depending on whether it is inherent or chosen. Furthermore, it appears to fluctuate depending on whether the referenced social identity aligns with the stance on the issue being discussed (e.g., extreme nationalist and pro-voter ID) or opposes it (e.g., democrats and pro-voter ID). To illustrate, in our Year 3 study, both social identities were aligned with the pro-voter ID standpoint of the tweets. In contrast, in our previous studies, the "minority" identity had been used to convey both pro- and anti-positions on issues such as the death penalty and voter ID requirements.

Univocality vs. Multivocality

During Year 1, our exploration underscored that while certain ideological groups focused steadfastly on a single issue across multiple platforms, others championed an array of issues, adapting their emphasis across platforms. Thus, Year 2 saw us delving into the effects of univocality (single-issue focus) and multivocality (multifaceted-issue focus) across platforms on consumers of ideological content. Our findings unveiled that univocality intensified users' emotional attachment to the primary issue, its personal significance, and perceptions of the content and its source's credibility. However, multivocality facilitated deeper information processing among users and their incorporation of the main issue in their responses to the content.

Violent Group Classification

In Year 2, our grouping of "violent" and "non-violent" factions was expanded in Year 3 through cluster analysis, further stratifying violent groups based on the nature and timing of their acts of violence, resulting in three distinct violent group clusters.

Foreshadowing Violence

Year 2 also encompassed an exploration of tweet content foreshadowing violence within domestic ideological groups. Our initial efforts using structural topic modeling identified religion as the most prevalent topic discussed pre-Jan. 6, 2021, accompanied by low emotive content. In Year 3, time series plots were constructed for both violent and non-violent groups surrounding the Jan. 6, 2021 event, to elucidate differences. Interestingly, both group types exhibited remarkable similarity in the 30 days leading up to and the 14 days following Jan. 6. Around Inauguration Day, however, non-violent groups displayed spikes in certain emotions and themes, while violent groups displayed slightly stronger relationship associations.

References

Connelly, S., Dunbar, N. E., Jensen, M. L., Griffith, J., Taylor, W. D., Johnson, G., ... & Mumford, M. D. (2015). Social categorization, moral disengagement, and credibility of ideological group websites. *Journal of Media Psychology*.

van der Vegt, I., Mozes, M., Kleinberg, B., & Gill, P. (2021). The grievance dictionary: Understanding threatening language use. *Behavior research methods*, 1-15.

Appendix A

Year 3 Final Report Definitions of Key Terms

Moral Disengagement	A cognitive mechanism by which one deactivates self-regulatory sanctions, allowing for rationalization, justification, or distancing from unethical or immoral behavior (Bandura, 1986). Consists of the cognitive mechanisms of moral justification, euphemistic labeling, advantageous comparison, diffusion and displacement of responsibility, distortion of consequences, dehumanization, and attribution of blame
Misinformation	Sharing and spreading of false, inaccurate, or misleading information
Time Series Analysis	Statistical technique used to evaluate data that has been collected over time to analyze and interpret patterns, trends, and relationships within the data as they evolve over time
Social Identities	A person's sense of who they are based on group membership(s) or feelings of connection to one or more groups.
Digital Trace Data	Information posted by extreme ideological groups on websites and microblogs

Appendix B

Dictionaries Used for Violent, Non-Violent Domestic Extremism Group Comparisons

Grievances - LIWC Dictionary Categories	Abbreviation	Example Identifiers
Desperation: It represents the presence of words related to feelings of hopelessness, helplessness, or extreme need.	desperation	hopelessness, despair, desperation, despondency, distress
Fixation: It represents the presence of words related to the act of obsessively focusing on a particular idea, object, or person.	fixation	obsession, preoccupation, infatuation, fixation, absorption
Frustration: It represents the presence of words related to feelings of annoyance, dissatisfaction, or being thwarted.	frustration	annoyance, irritation, exasperation, vexation, dissatisfaction
God: It represents the presence of words related to the concept of a higher power or deity.	god	deity, divinity, higher power, supreme being, religious figure
Grievance: It represents the presence of words related to a complaint or feeling of injustice.	grievance	complaint, injustice, grievance, resentment, dissatisfaction
Hate: It represents the presence of words related to intense dislike, hostility, or animosity.	hate	dislike, animosity, loathing, hatred, antipathy
Help: It represents the presence of words related to seeking assistance, support, or aid.	help	aid, support, assistance, rescue, relief
Honour: It represents the presence of words related to the concept of respect, integrity, or esteem.	honour	integrity, dignity, respect, virtue, esteem
Impostor: It represents the presence of words related to someone pretending to be someone else or deceiving others.	impostor	fraud, fake, pretender, impostor, charlatan
Jealousy: It represents the presence of words related to the feeling of envy or resentment toward someone else's possessions, qualities, or achievements.	jealousy	envy, resentment, covetousness, bitterness, rivalry
Loneliness: It represents the presence of words related to the state of being alone or feeling socially isolated.	loneliness	isolation, solitude, desolation, abandonment, alienation

Murder: It represents the presence of words related to the act of unlawfully killing another person.	murder	kill, homicide, assassination, slay, manslaughter
Paranoia: It represents the presence of words related to the feeling of extreme distrust or suspicion of others.	paranoia	suspicion, distrust, obsession, delusion, fear
Planning: It represents the presence of words related to the act of making detailed arrangements or strategies for a particular purpose.	planning	strategizing, scheming, plotting, preparing, organizing
Relationship: It represents the presence of words related to connections, associations, or interactions between individuals.	relationship	connection, bond, partnership, association, rapport
Soldier: It represents the presence of words related to military personnel or the act of serving in the armed forces.	soldier	warrior, fighter, serviceman, combatant, trooper
Suicide: It represents the presence of words related to the act of intentionally causing one's own death.	suicide	self-harm, self-destruction, suicidal, despair, hopelessness
Surveillance: It represents the presence of words related to monitoring or observing others, often with the intention of gathering information.	surveillance	monitoring, observation, scrutiny, spying, tracking
Threat: It represents the presence of words related to expressing intent or potential to cause harm, danger, or damage.	threat	danger, menace, intimidation, peril, warning
Violence: It represents the presence of words related to the use of physical force or aggression to cause harm or damage.	violence	brutality, aggression, force, cruelty, brutality
Weaponry: It represents the presence of words related to weapons or tools used for combat or defense.	weaponry	arms, firearms, munitions, arsenal, weaponry
Hate speech: It represents the presence of words or language that promotes or incites hatred, discrimination, or violence toward a particular group.	hatespeech	bigotry, discrimination, prejudice, intolerance, racism
Hate speech2: This likely refers to an additional category or subcategory of hate speech.	hatespeech2	homophobia, xenophobia, sexism, hate speech, bias
Hate speech3: This likely refers to another category or subcategory of hate speech.	hatespeech3	Islamophobia, antisemitism, transphobia, hate crime, harassment

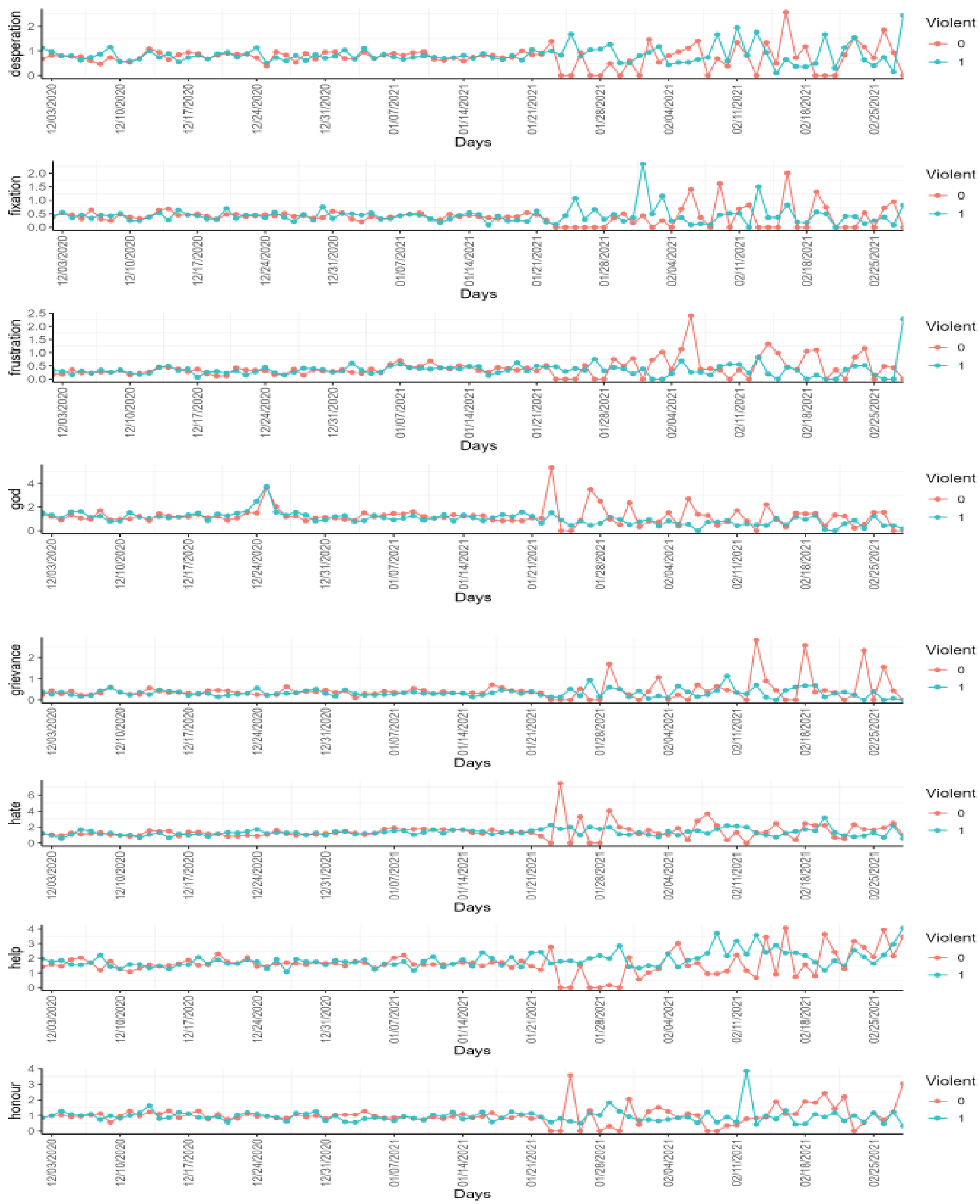
Plutchik Emotion Categories:	Example Words (0 Low/no intensity 10 High intensity)
Negative Emotion – represents emotions associated with unpleasant or adverse experiences, feelings, or states.	• 0 – neutral, calm, content, indifferent • 10 – anger, rage, fear, despair, grief, hatred
Positive Emotion – represents emotions associated with pleasant or positive experiences, feelings, or states.	• 0 – neutral, satisfied, content, pleased • 10 – joy, ecstasy, excitement, elation, love, happiness

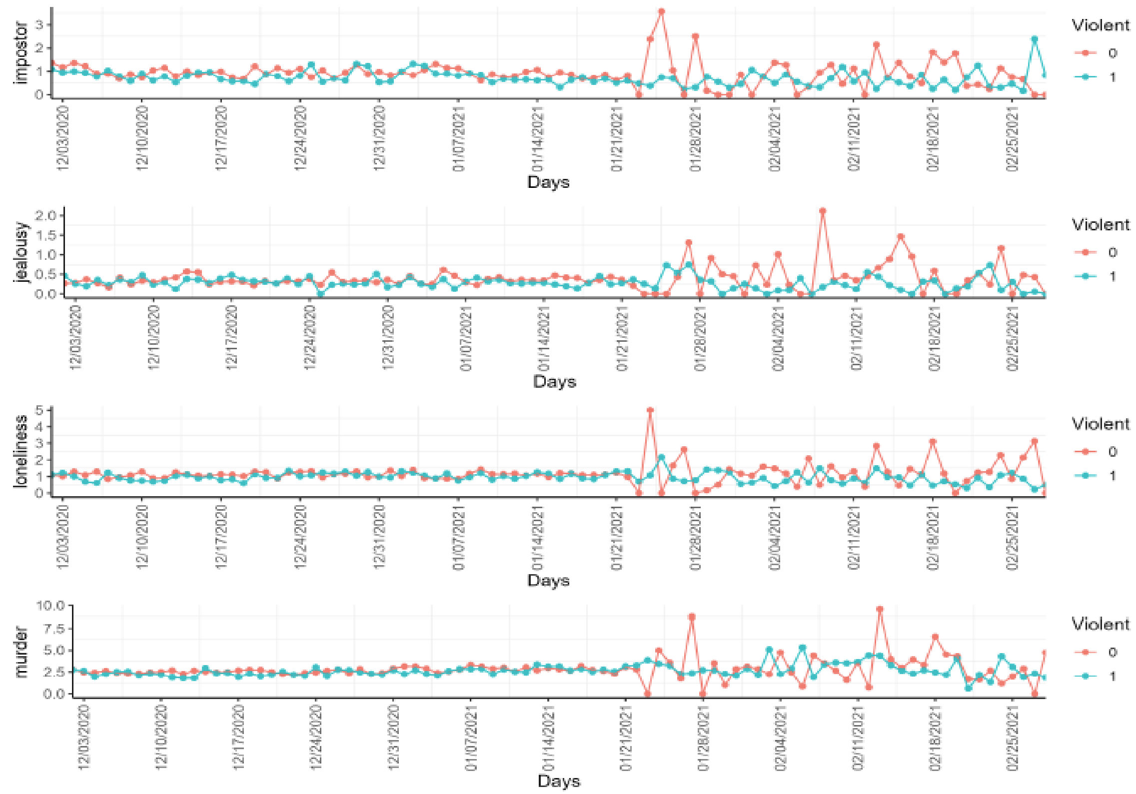
LIWC 22 - Dictionary Categories:	Abbreviation	Example Identifiers
Positive Tone: Represents words with positive emotional connotations.	tone_pos	pos: positive tone, positive sentiment, positive attitude
Negative Tone: Represents words with negative emotional connotations.	tone_neg	negative tone, negative sentiment, negative attitude
Positive Emotion: Represents words associated with positive emotions.	emo_pos	positive emotions, positive feelings, positive affect
Negative Emotion: Represents words associated with negative emotions.	emo_neg	negative emotions, negative feelings, negative affect
Insight and introspection. Represents words associated with self-reflection, understanding, or gaining insights.	insight	insight, understanding, realization, perception
Causal words. Indicates the presence of words related to cause-and-effect relationships or explanations.	cause	cause, reason, motive, factor
Discrepancy. Represents words related to inconsistencies, contradictions, or variations.	discrep	discrepancy, inconsistency, difference, variation
Tentative words. Indicates the presence of words associated with uncertainty, hesitation, or cautiousness.	tentat	tentative, uncertain, hesitant, doubtful
Certainty words. Represents words related to confidence, certainty, or strong beliefs.	certitude	certainty, confidence, assurance, conviction
Differences. Indicates the presence of words related to distinctions, variations, or disagreements.	differ	differences, different, diverse, vary

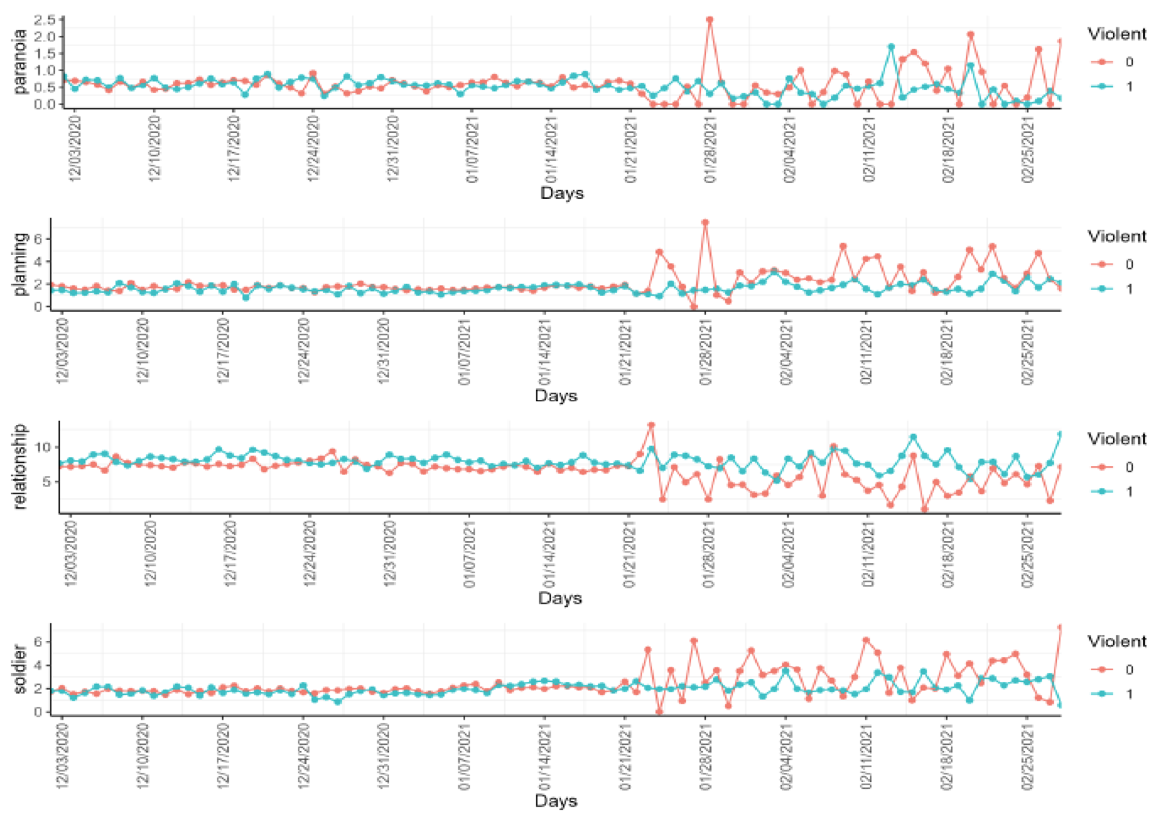
Affiliation. Indicates the presence of words related to social connections, group membership, or affiliation.	affiliation	affiliation, connection, belonging, membership
Achievement. Represents words associated with accomplishment, success, or reaching goals.	achieve	achievement, success, accomplish, attain
Power. Indicates the presence of words related to power, influence, or control.	power	power, influence, control, authority

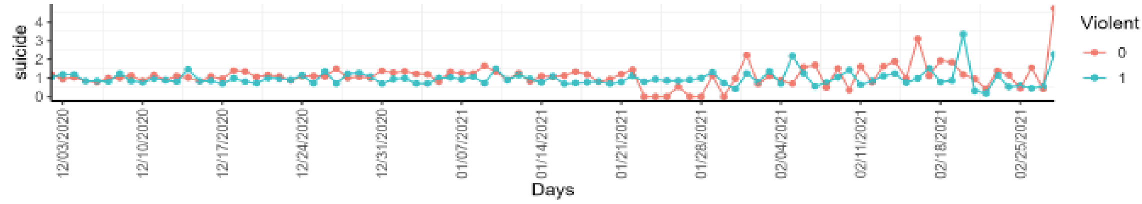
Appendix C

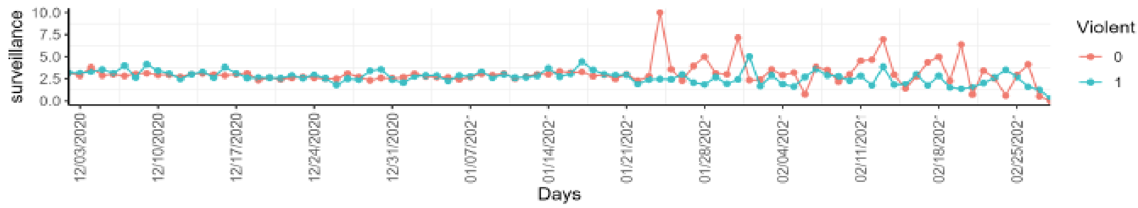
Time Series Graphs for Violent and Non-violent Group Tweets Before and After the January 6th 2021 Attack on the U.S. Capital Building

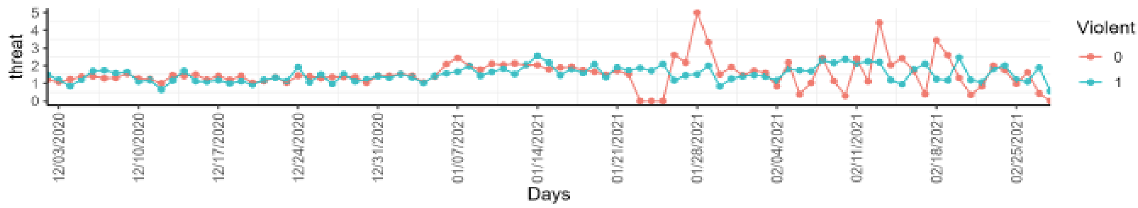


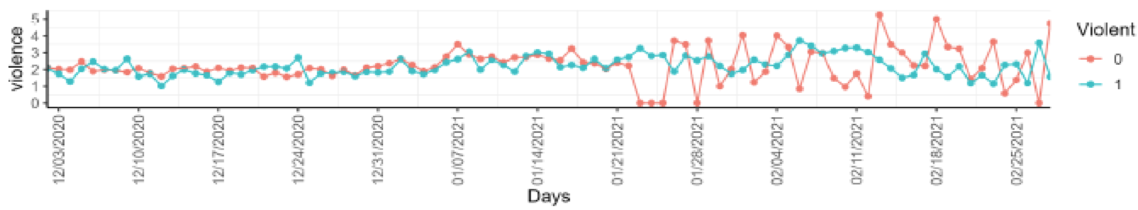

impostor
jealousy
loneliness
murder
Days
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Violent
0
1

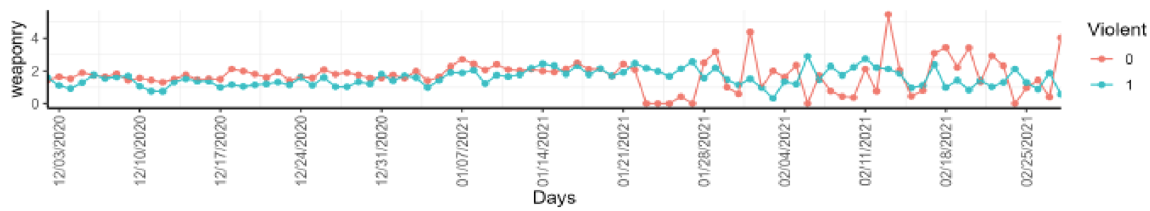

paranoia
planning
relationship
soldier
Days
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Violent
0
1

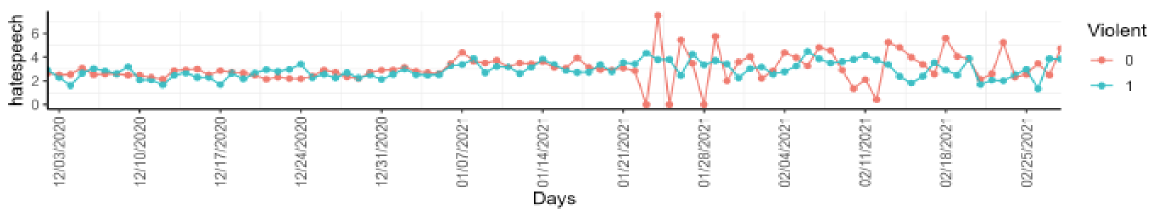

suicide
0
1
2
3
4
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

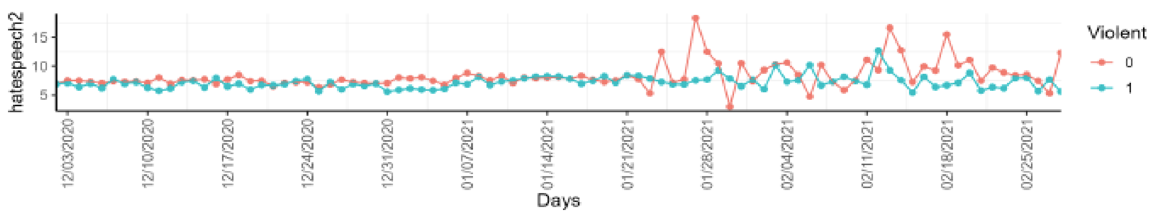

surveillance
0.0
2.5
5.0
7.5
10.0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
Days
Violent
0
1

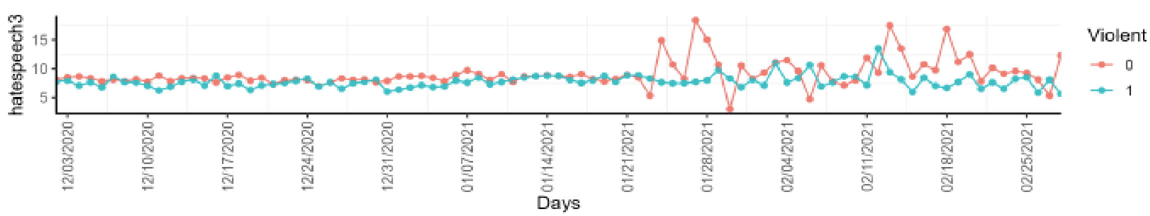

threat
0
1
2
3
4
5
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

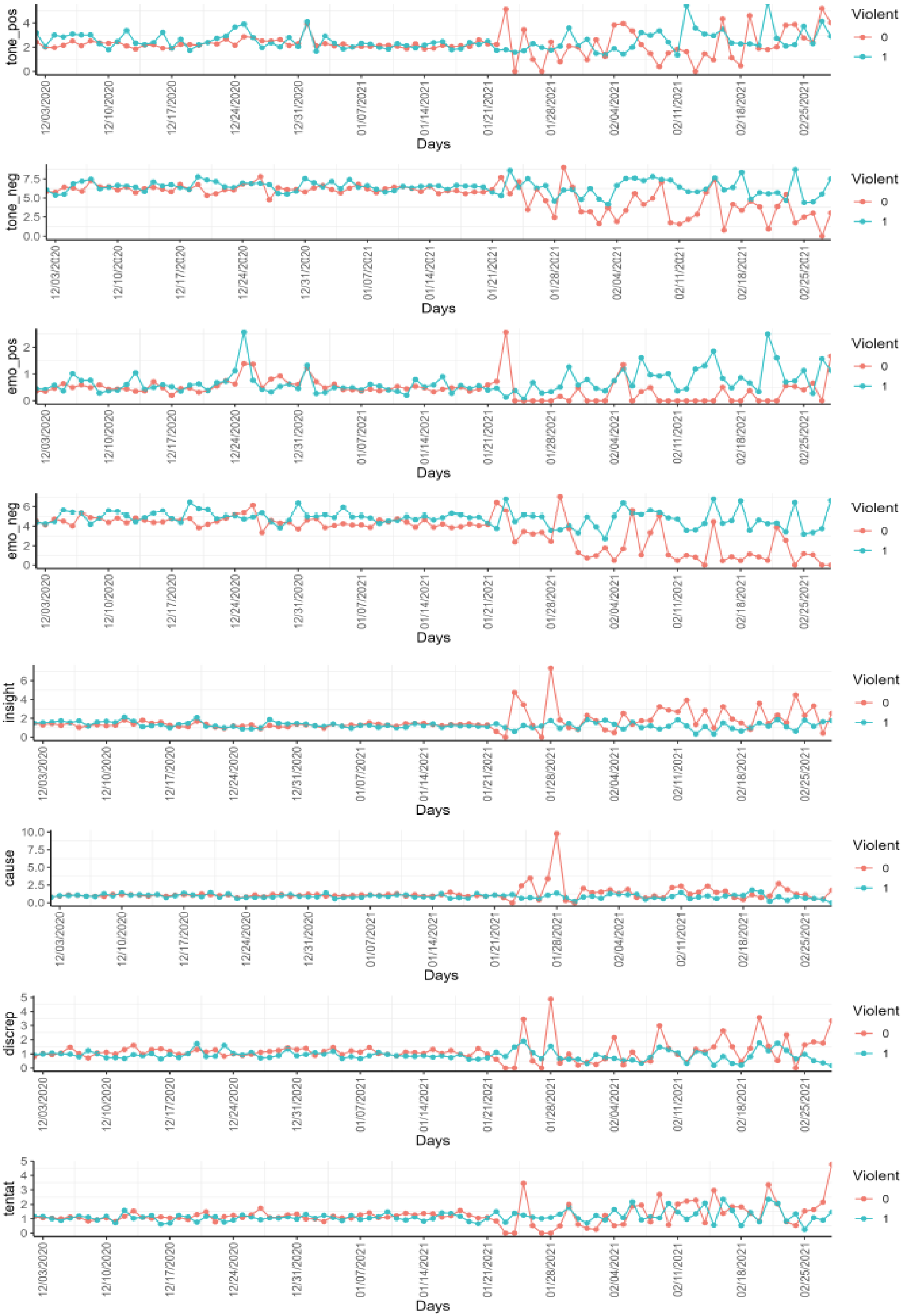

violence
0
1
2
3
4
5
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

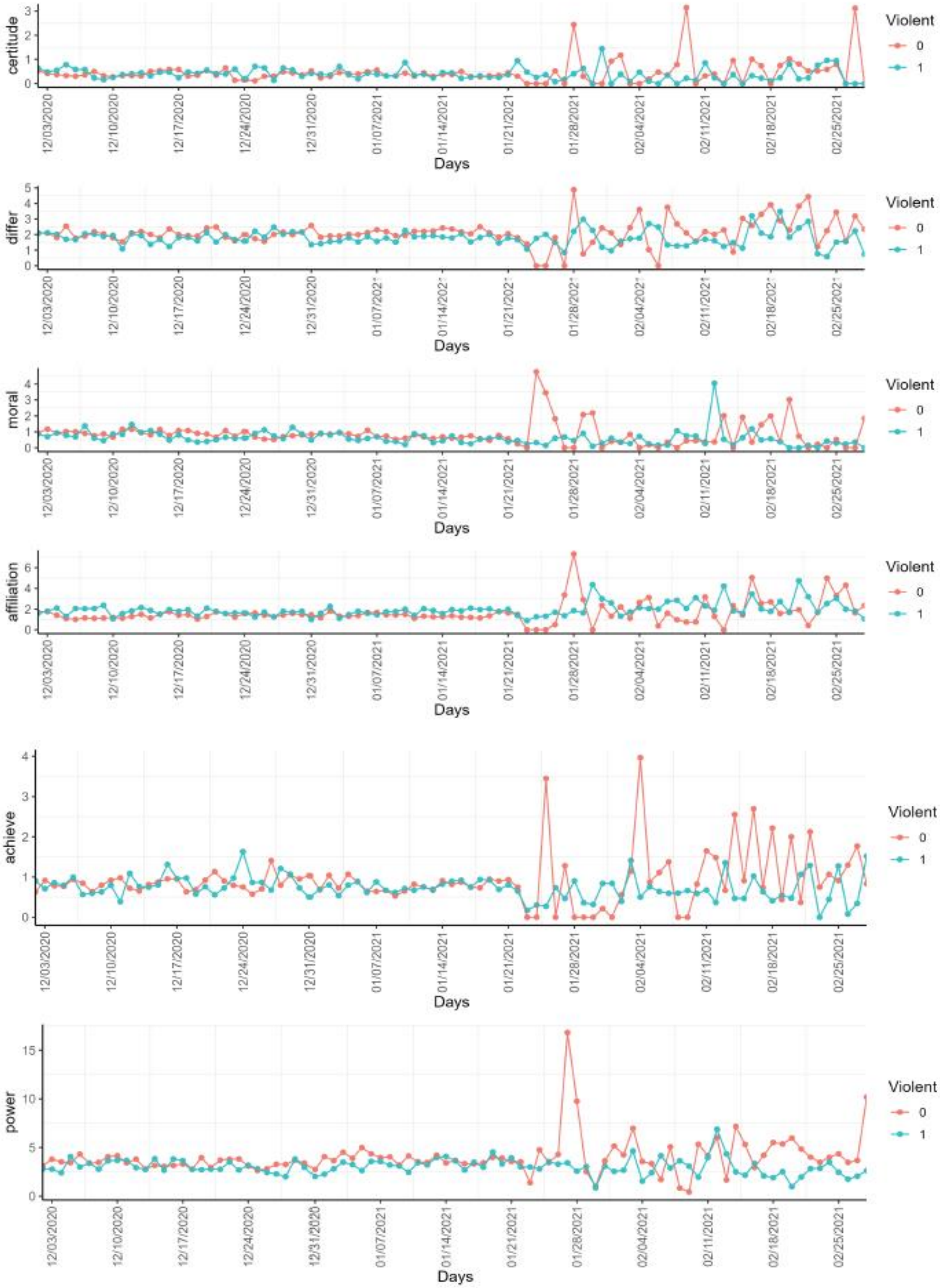

weaponry
0
2
4
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

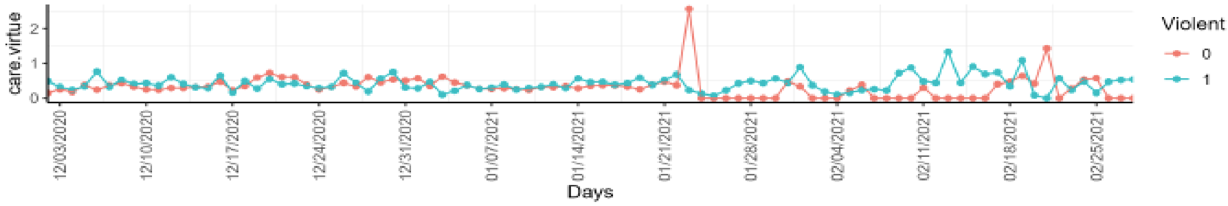

hatespeech
0
2
4
6
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
Days
Violent
0
1

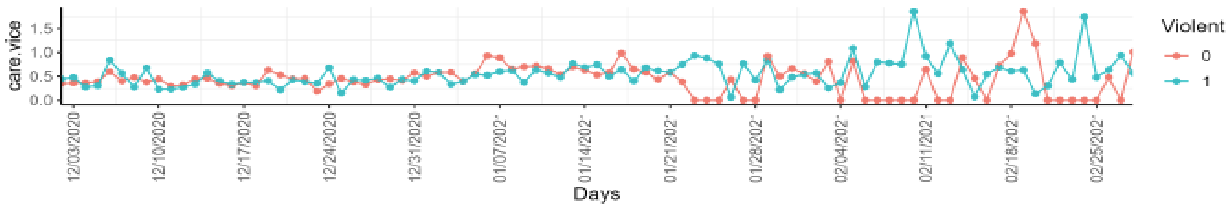

hatespeech2
5
10
15
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

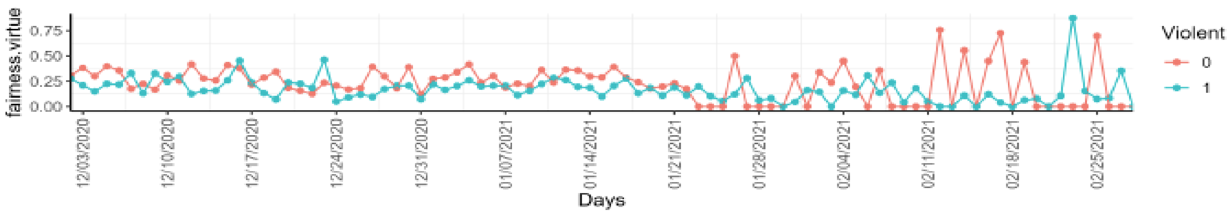

hatespeech3
5
10
15
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

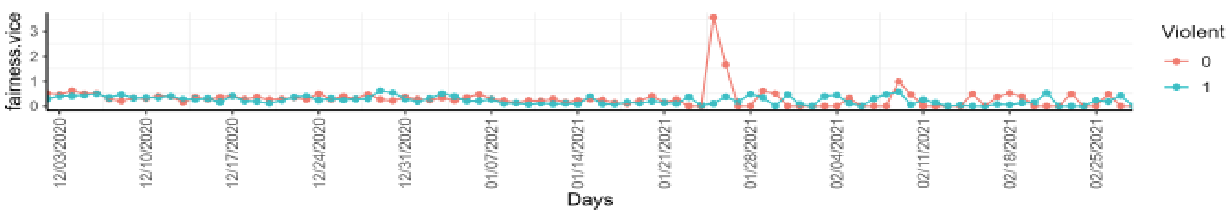

tone_pos
Days
Violent
0
1
tone_neg
Days
Violent
0
1
emo_pos
Days
Violent
0
1
emo_neg
Days
Violent
0
1
insight
Days
Violent
0
1
cause
Days
Violent
0
1
discrep
Days
Violent
0
1
tentat
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

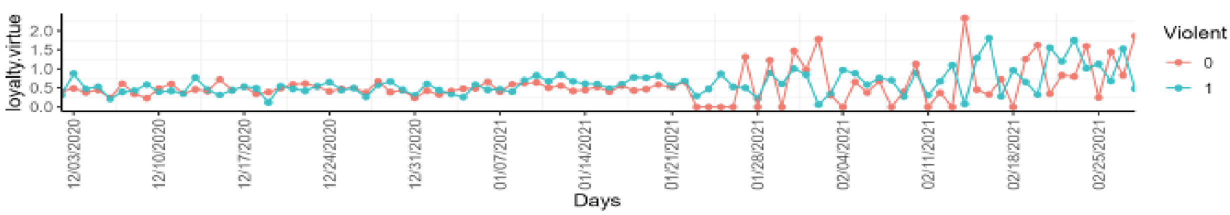

certitude
differ
moral
affiliation
achieve
power
Days
Violent
0
1
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021

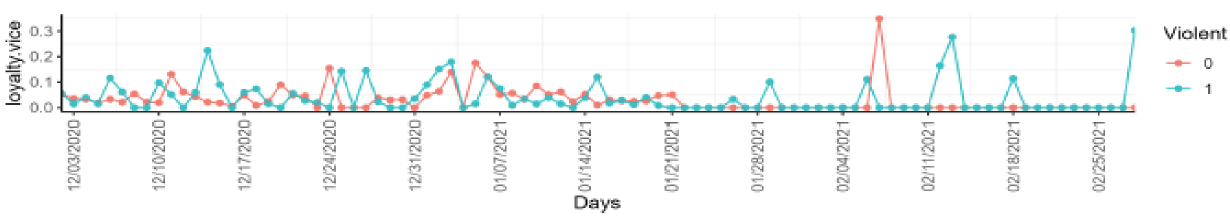

care.virtue
0
1
2
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

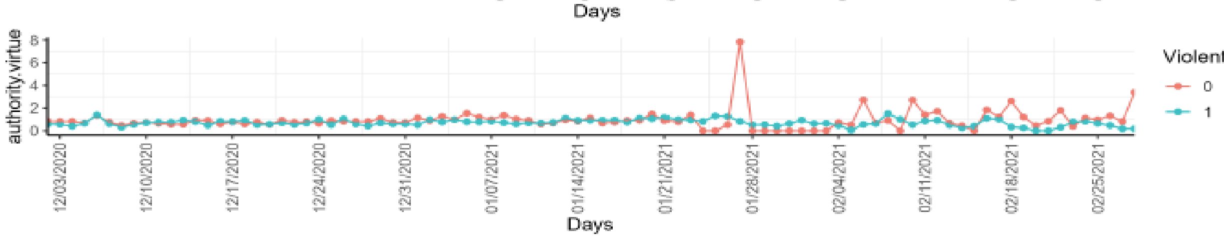

care.vice
0.0
0.5
1.0
1.5
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/202
01/14/202
01/21/202
01/28/202
02/04/202
02/11/202
02/18/202
02/25/202
Days
Violent
0
1

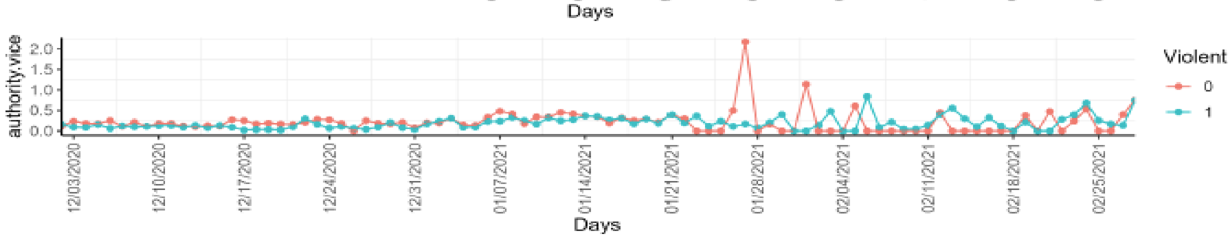

fairness.virtue
0.00
0.25
0.50
0.75
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

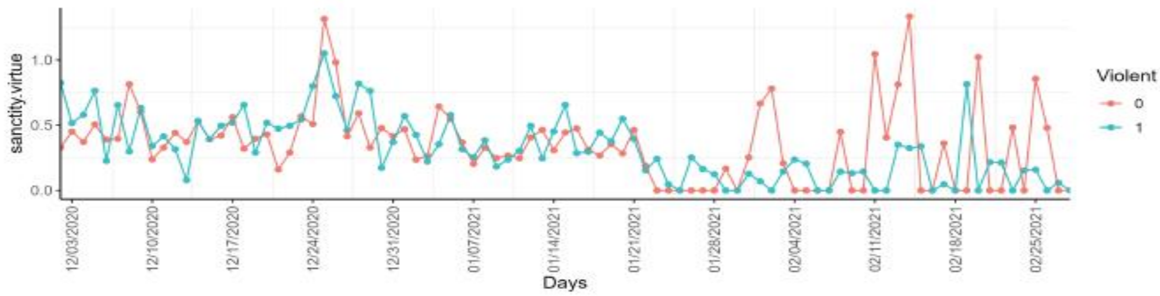

fairness.vice
0
1
2
3
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1


loyalty.virtue
0.0
0.5
1.0
1.5
2.0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1

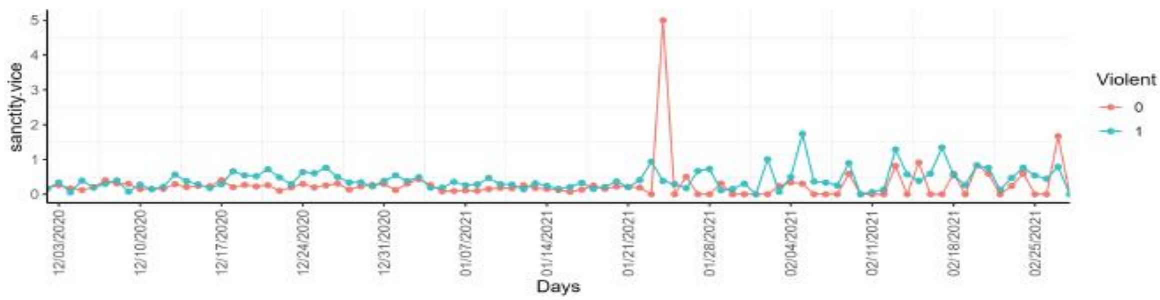

loyalty.vice
0.0
0.1
0.2
0.3
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1


authority.virtue
0
2
4
6
8
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1


authority.vice
0.0
0.5
1.0
1.5
2.0
12/03/2020
12/10/2020
12/17/2020
12/24/2020
12/31/2020
01/07/2021
01/14/2021
01/21/2021
01/28/2021
02/04/2021
02/11/2021
02/18/2021
02/25/2021
Days
Violent
0
1



sanctity.virtue
Days
Violent
0
1



sanctity.vice
Days
Violent
0
1