American Political Science Review (2024) 1–18

doi:10.1017/S0003055423001478

"Keep Your Heads Held High Boys!": Examining the Relationship between the Proud Boys' Online Discourse and Offline Activities

CATIE SNOW BAILARD *George Washington University, United States*
REBEKAH TROMBLE *George Washington University, United States*
WEI ZHONG *New York University, United States*
FEDERICO BIANCHI *Stanford University, United States*
PEDRAM HOSSEINI *George Washington University, United States*
DAVID BRONIATOWSKI *George Washington University, United States*

This study examines the relationship between online communication by the Proud Boys and their offline activities. We use a supervised machine learning model to analyze a novel dataset of Proud Boys Telegram messages, merged with US Crisis Monitor data of violent and nonviolent events in which group members participated over a 31-month period. Our analysis finds that intensifying expressions of grievances online predict participation in offline violence, whereas motivational appeals to group pride, morale, or solidarity share a reciprocal relationship with participation in offline events. This suggests a potential online messaging–offline action cycle, in which (a) nonviolent offline protests predict an increasing proportion of motivational messaging and (b) increases in the frequency and proportion of motivational appeals online, in turn, predict subsequent violent offline activities. Our findings offer useful theoretical insights for understanding the relationship between online speech and offline behavior.

INTRODUCTION

While we know that a number of far-right groups' offline activities—including the Stop the Steal rally and subsequent insurrection at the US Capitol on January 6, 2021—have been organized, in part, online, it remains difficult to assess whether and in what manner routine, everyday online discussions among right-wing extremist groups' adherents relate to their offline behavior. To better understand this relationship, we must overcome at least two interrelated challenges. The first challenge is technical. Analyzing the incredible volume of online communication has proven difficult (Vidgen et al. 2019), even for the largest and best-resourced social media platforms. And yet, we argue, the more fundamental challenge is a theoretical one. To enhance our analytical capabilities, analyses of online content—even large-scale online content—must draw on and develop technical approaches that are grounded in the rich, nuanced, and more complex social scientific theories of political communication and mobilization. In other words, by viewing online content not as a set of individual posts containing discrete messaging, but rather as a complex and rich discursive environment, in which messages ebb and flow over time, we are more likely to uncover the mobilizing power and potential of online communication.

In this study, we examine the relationship between the online communication of one prominent right-wing extremist group—the Proud Boys—and the group's offline activities using the long-standing and well-developed collective action framing literature as the core theoretical lens driving our analytical approach. The power of messages and speech to motivate and mobilize group members lies at the heart of collective action framing theory. According to the foundational work of Goffman (1974) and Snow and Benford (1988), collective action frames denote "schemata of interpretation" that enable individuals "to locate, perceive, identify, and label" occurrences within their life space and the world at large (Benford and Snow 2000, 21). Frames help to render events or occurrences meaningful and thereby function to organize experience and guide action.

Corresponding author: Catie Snow Bailard, Associate Professor, School of Media and Public Affairs, George Washington University, United States, cbailard@gwu.edu.

Rebekah Tromble, Associate Professor, School of Media and Public Affairs, George Washington University, United States, rtromble@gwu.edu.

Wei Zhong, Postdoctoral Fellow, Center for Social Media and Politics, New York University, United States, wz490@nyu.edu.

Federico Bianchi, Postdoctoral Researcher, Department of Computer Science, Stanford University, United States, fede@stanford.edu.

Pedram Hosseini, Alumnus, School of Engineering and Applied Science, George Washington University, United States, phosseini@gwu.edu.

David Broniatowski, Associate Professor, School of Engineering and Applied Science, George Washington University, United States, broniatowski@gwu.edu.

Received: September 09, 2022; revised: March 29, 2023; accepted: December 20, 2023.

Collective action frames simplify and condense aspects of the "world out there," but in ways that are "intended to mobilize potential adherents and constituents, to garner bystander support, and to demobilize antagonists" (Benford and Snow 2000, 198). This framework has been used by scholars to examine the communication patterns of a multitude of violent and nonviolent social movement organizations. In particular, research has explored the ways in which social movement actors deploy collective action frames to promote group identity and solidarity, motivate actions, and interpret and make sense of information that could challenge the group's foundational beliefs (Bos et al. 2020; Goh and Pang 2016; Oktavianus, Davidson, and Guan 2021).

Following this tradition, our analysis focuses on the degree to which members of the Proud Boys deploy three high-level categories of collective action frames—diagnostic, prognostic, and motivational—and two subcategories of diagnostic frames—injustice and othering—in their everyday conversations on the social media platform Telegram (Benford and Snow 2000; Hunt, Benford, and Snow 1994; Snow and Benford 1988). We then examine the temporal correlation of these frames with both violent and nonviolent offline activities by members of this extremist group.

Carefully operationalizing these concepts, we finetune a state-of-the-art supervised machine learning model that allows us to identify each type of frame and subframe at scale and analyze their variation over time. Applying the models to more than 500,000 messages posted on 92 Proud Boys-affiliated Telegram channels between January 1, 2020, and July 26, 2022, our empirical analysis reveals that diagnostic messages—those that identify a problem and/or assign attribution for a problem perceived to afflict the group—are effectively omnipresent in online discussions. However, increases in the proportion of these messages predict subsequent offline violence. Motivational messages—those that boost in-group morale, pride, or solidarity—on the other hand, prove both responsive to and predictive of offline activities. As our analysis suggests, these solidarity-building messages tend to increase in proportion after nonviolent offline protests. In turn, upticks in the proportion and frequency of motivational frames prove a strong predictor of offline violence. We refer to these reciprocal dynamics as the *online messaging–offline action cycle.*

These findings empirically support the capacity for collective action frames to mobilize group members, as posited by the social movement literature. However, if we wish to understand the cyclical dynamics observed in our analysis, we argue that we must go beyond the social movement framework, supplementing it with insights from research on "moralizing" and "moral convergence" (Mooijman et al. 2018). Diagnostic collective action frames, we suggest, lay the groundwork for offline mobilization by generating moral justification for action. Diagnostic frames are foundational to an extremist group's sense of their place in an unjust world. However, in addition to believing that action is morally justified, engaging in violence is more likely when members of extremist organizations also trust that they are acting *in common cause*, on the basis of shared values. Prompted by offline encounters that enhance this sensibility, solidarity-building motivational frames rise in prominence online, and the prominence of these messages may in turn help to spur offline violence.

In laying out these arguments and findings, we begin with a discussion of the core theoretical insights guiding our analysis. Next, we turn to our research design and methodologies, including an overview of our case study, the Proud Boys' communication on Telegram. This section also offers a description of our computational approaches and how they relate to previous work and ends with an overview of our statistical analyses. We then lay out our findings and conclude with a discussion of their theoretical and practical implications.

THEORETICAL FRAMEWORK

Online Communication and Offline Activities

Relative to previous eras, social media has empowered a more diffuse set of grassroots actors and groups to organize and mobilize its members for action (Shirky 2008). A growing body of research has substantiated the correlation between Internet use and political participation (see Boulianne 2015), illuminating a number of dynamics and nuances in this relationship. For example, one study found that specific types of social media use (e.g., expressive, informational, and relational) increased citizen engagement, whereas other uses (e.g., identity and entertainment) did not (Skoric et al. 2016). Other research has shown that the affordances of different platforms affect the likelihood of participation in distinct ways. For instance, while strong ties on Facebook prove more effective for protest-related communication, weak ties appear to be key on Twitter (Valenzuela, Correa, and Gil de Zúñiga 2018).

A large body of research has also documented the ways in which right-wing extremist groups specifically make use of the Internet to recruit followers, broadcast their intolerant worldview and rhetoric, inculcate a sense of identity and community among members, and organize offline activities (Greene 2019; Munn 2019; Scrivens, Davies, and Frank 2020). However, the body of systematic empirical inquiries into the correlation between online speech and offline activities by right-wing extremist groups is still nascent (Müller and Schwarz 2021; 2023; Williams et al. 2020). One analysis of word frequencies in Proud Boys' Gab communications before and after the far-right "Free Speech" rally in Berkeley and the Unite the Right riot in Charlottesville found that post-event Gab posts included more incident-driven rhetoric than the pre-rally speech, which the researchers suggest may serve a myth-making function for the group in the wake of controversial activities (Reid, Valasik, and Bagavathi 2020). Another study found that increased online engagement on Facebook between oppositional protest groups predicted increased violence between these groups when they met offline (Gallacher, Heerdink, and Hewstone 2021).

Online Communication and Social Identity

Another emergent body of literature applies concepts related to group identity to analyses of social media data. Much of this work draws from social identity theory (Tajfel and Turner 2004), in which individuals' sense of self-worth, belongingness, and identity largely stem from their membership in groups. Moreover, the in-groups to which an individual belongs are often defined and evaluated relative to other groups (i.e., out-groups). The resulting tension shapes many dimensions of intergroup interactions and conflict. An analysis of the recent resurgence of populism, for example, demonstrates how in-group/out-group frames that pit the common man against corrupt elites for governmental failings increase individuals' support for populist candidates and ideas (Busby, Gubler, and Hawkins 2019).

Social media platforms offer a low-cost, accessible, relatively efficient platform for group members to find one another to coalesce around and negotiate their collective identity and perceived position in society relative to other groups, as well as organize for collective action on behalf of the group's interests (Khazraee and Novak 2018; Makki et al. 2018; Velasquez and Montgomery 2020; Wang, Liu, and Gao 2016). "Social groupings and divisions are constituted in and through communication…the identities that are consequential for politics are not structural givens, they are created over time by political and social actors and made salient at particular moments as an organizing basis of political life" (Kreiss, Lawrence, and McGregor 2020, 4). For example, a recent analysis of Facebook and Twitter has found that posts mentioning a political out-group were shared approximately twice as often as those about the in-group (Rathje, Van Bavel, and Van 2021). Within the context of right-wing extremism, there is also evidence of white supremacist groups using memes for socialization and identity-building (DeCook 2018), and another study shows how white nationalists use the online message board Stormfront to attempt to appeal to mainstream whites, by emphasizing pride and communal well-being that creates a rhetorical distance between white nationalism and white supremacy (Hartzell 2020).

Collective Action Framing

Pulling these strands of literature together, our analysis considers whether and how online communication that features specific types of collective action frames might increase the likelihood that members of a right-wing extremist group will participate in offline events. We focus on collective action frames because these often serve to construct and maintain group identity, create a shared worldview and understanding of where that group fits into society, and mobilize members to act on behalf of the group. Similar to media frames discussed in the political communication literature, collective action frames highlight specific dimensions of an issue or event and call to mind certain values, allowing the frames' purveyors—whether media outlets or social movement organizations—the ability to influence how their audience perceives and interprets the information relayed in a message and, in some instances, shaping their reactions to it (Chong and Druckman 2007). Whereas broadcast media traditionally had substantial latitude to select and deploy such frames, social media has empowered average citizens and grassroots groups to, on occasion, challenge and recast even dominant frames (Guggenheim et al. 2015; Jackson and Foucault 2015).

In this study, we investigate the volume and proportion of Telegram messages posted to Proud Boys-affiliated channels that feature one or more of three high-level categories of collective action frames—diagnostic, prognostic, and motivational—as well as two subcategories of diagnostic frames—injustice and othering. (Please see Supplementary Table A2 for more information regarding the annotation scheme and examples of each of the frames.)

Diagnostic Collective Action Frames

We begin with diagnostic frames, which identify some problem and/or attribute blame for a problem (Benford and Snow 2000) that is perceived to negatively affect the group and its members. For example, right-wing extremist groups commonly decry censorship, perceived threats to their racial or religious identity, the decline of certain social values or norms, and societal changes that they believe have emasculated or diminished men's traditional role in society. Within this diagnostic frame category, we examine two prevalent subcategories—namely, "injustice" and "othering" frames. These subframes are identified in relevant bodies of literature as both prevalent components of right-wing extremist discourse and effective collective action frames for mobilizing members of groups more broadly.

Injustice frames define an action, policy, event, or phenomenon as fundamentally unfair to those within an in-group. Previous studies substantiate the prevalence of injustice frames in the rhetoric used by white extremist groups (Adams and Roscigno 2005; Berbrier 1998; 2000; Bubolz and Simi 2019), in which a Manichean worldview portrays whites as the aggrieved in-group. When a group's identity centers on a shared sense of being aggrieved, this increases the likelihood of mobilization because members are more likely to engage in collective action when its goal is to regain their in-group's "appropriate" or "deserved" status by confronting those they believe are responsible for this state of affairs (Kawakami and Dion 1995; Van Zomeren, Postmes, and Spears 2008). Moreover, when victimhood and injustice are associated with group identity, members are more likely to feel anger about that perceived injustice, which increases the likelihood that members will take action on behalf of the group (Mackie, Devos, and Smith 2000).

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

Whereas injustice frames focus on the perceived unjust treatment of the in-group, othering frames turn the focus to out-groups. Increasing the perceived threat an out-group poses, dehumanizing its members, and/or amplifying the perceived incompatibility between the groups and their ways of life can be an effective catalyst for mobilizing in-group members to act against an out-group (Postmes et al. 1999; Simon and Klandermans 2001; Van Zomeren, Postmes, and Spears 2008). "Research in the field of identity framing has indicated that in-group mobilization results from priming a severe threat to the well-being of the group… motivating the in-group to take action" (Bos et al. 2020, 6). This amplified sense of threat also helps provide moral cover or justification for the actions taken to defend the in-group from this threatening other. "In concrete terms, this could be manifest as the argument that white supremacists are simply concerned with the survival of their people, and that if some on the fringe feel that urgent action is required as a result of dangers posed by sinister outside forces, that is understandable" (Berbrier 2000, 187).

Taken together, diagnostic frames generate a sense of threat against and, crucially, a moral justification for acting to protect (Mooijman et al. 2018; Skitka, Bauman, and Sargis 2005; Skitka and Morgan 2014) group members and their interests. This leads to the following hypotheses:

Hypothesis 1: *An increase in the frequency and/or proportion of diagnostic frames in online communications among members of the Proud Boys predicts an increased likelihood of its members participating in an offline event.*

Hypothesis 1A: *An increase in the frequency and/or proportion of injustice frames in online communications among members of the Proud Boys predicts an increased likelihood of its members participating in an offline event.*

Hypothesis 1B: *An increase in the frequency and/or proportion of othering frames in online communications among members of the Proud Boys predicts an increased likelihood of its members participating in an offline event.*

Prognostic Collective Action Frames

If diagnostic frames point to problems and attribute blame for those problems, prognostic frames, in turn, "suggest solutions" and "identify strategies, tactics, and targets" for addressing a problem (Benford and Snow 2000, 201). The logic connecting prognostic frames with offline participation is perhaps the most self-evident, since by definition these messages entail some sort of call-to-action to protest, address, or take action to ameliorate some perceived problem afflicting the in-group. However, in some cases, the solution presented is more abstract and broadly conceived, including calls to come together to "do something" to reclaim the group's perceived way of life.

In considering prognostic frames, we hypothesize as follows:

Hypothesis 2: *An increase in the frequency and/or proportion of prognostic frames in online communications among members of the Proud Boys predicts an increased likelihood of its members participating in an offline event.*

Motivational Collective Action Frames

Lastly, motivational frames provide "a rationale for action" or a "vocabulary of motives" for mobilizing (Snow and Benford 1988, 202) by positively priming some aspect of the in-group identity. Motivational frames serve to boost morale, pride, and/or a sense of belonging or duty to that group, each of which increases the likelihood of mobilization (Benford and Snow 2000; Berbrier 1998; 2000; Van Zomeren, Postmes, and Spears 2008). To this end, these messages often reference shared values, principles, priorities, norms, and/or characteristics of that group's identity. Additionally, motivational messages can serve to increase the group's sense of its own strength, numbers, and/or support from powerful allies, tilting the cost–benefit calculus in favor of engaging in an offline activity. "Appraisals of in-group strength produce anger toward an opponent out-group, and anger is a potent predictor of offensive action tendencies" (Mackie, Devos, and Smith 2000, 613; see also Mooijman et al. 2018; Skitka, Bauman, and Sargis 2005), which is exemplified by the sharp increase in hate crimes and extremist activities in the years following the election of Donald Trump to the presidency (Williamson and Gelfand 2019).

Finally, motivational frames can "supply adherents with compelling reasons or rationales for taking action and provide participants with justifications for actions undertaken on behalf of the movement's goals, particularly when their behavior is called into question by friends, family or coworkers" (Benford and Hunt 1992, 41). In this sense, these frames can also function to destigmatize, mainstream, and legitimize the group's image and intolerant worldview. For right-wing extremists, "The mandate to love and have pride in one's heritage are presented as universally valid… People are exhorted to 'love your heritage and love your culture' and 'everyone should be allowed to do so'" (Berbrier 1998, 442). Such messages can, thus, also function to provide moral cover for actions taken by group members.

This generates the following prediction:

Hypothesis 3: *An increase in the frequency and/or proportion of motivational frames in online communications among members of the Proud Boys predicts an increased likelihood of its members participating in an offline event.*

RESEARCH DESIGN AND METHODOLOGY

Case Study—The Proud Boys and Telegram

Founded by far-right political commentator Gavin McInnes in 2016, the Proud Boys are an all-male, far-right neo-fascist group that describes themselves as "proud Western chauvinists who refuse to apologize for creating the modern world" (McBain 2020). Classified as a far-right extremist organization by the FBI

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

(Wilson 2018), as of 2021, they had 116 chapters across 46 states. Proud Boys members often mobilize alongside other white supremacist groups, including at high-profile events such as the Charlottesville Unite the Right rally and the UC Berkeley protests of 2017. The Proud Boys also regularly participate in violent street brawls and frequently use social media to encourage violence toward perceived enemies (Anti-Defamation League 2018). Their visible and documented participation in the January 6 insurrection has also made them a focal point for prosecution and investigation efforts by multiple government agencies (Walker 2022).

The Proud Boys have long relied on a variety of social media platforms for both internal and external communication—recruiting new members, planning activities, growing solidarity among their adherents, and attempting to share and mainstream their beliefs with the wider public. However, in recent years, major social media platforms have taken relatively strict enforcement action against the Proud Boys, including the founder McInnes. In 2018, both Twitter and Facebook announced that they would deplatform the group and quickly removed the largest accounts linked to the organization and its leaders (Linton 2018). Like other right-wing extremist groups, the Proud Boys responded to these bans by migrating to smaller, alternative social media platforms, such as Telegram. As of 2021, there was a multitude of Proud Boys Telegram channels, the largest of which had more than 28,000 members.

Telegram is a free cloud-based instant messaging platform created in 2013 by brothers Pavel and Nikolai Durov, amid their own troubles with the authoritarian Russian state's censorship of online speech (Urman and Katz 2022). Telegram offers several options for engagement, including private one-on-one conversations, group chats, and both private and public channels that are controlled by admins. According to the Telegram website,[1] the company does not partake in extensive content-moderation policies, only removing pornographic material and blocking ISIS-related terrorist activity on its public channels (Walther and McCoy 2021). As a result, Telegram has attracted users from across a range of hate-based organizations.

Though Telegram has been a favored platform for extremist groups for a number of years—with its popularity increasing dramatically following the January 6th Capitol insurrection (Dickson 2021)—to date, relatively few studies have been published that investigate how these groups make use of Telegram. One recent analysis of the content of a dozen far-right Telegram channels identified white grievances, and blaming minorities for those grievances, as dominant themes in their discussions (Al-Rawi 2021)—paralleling the diagnostic subframes of injustice and othering that we investigate here. Another study of three far-right groups in Germany (QAnon, Identitarian Movement, and Querdenken) identified the presence of several radicalizing narratives, including anti-elitism, support for violence, activism, and conspiracy theories (Schulze et al. 2022)—types of speech that have implications for mobilization.

Data Collection

We collected our data from publicly viewable Proud Boys-affiliated Telegram channels. Typically, each Telegram channel is set up by its owner to allow broadcast or one-way communication from a small set of senders to a broader set of general channel users. However, some channels are interactive, allowing channel followers to share and respond to posts. Content in a channel primarily consists of messages comprised of text, images, audio files, and/or video files. To collect content and metadata from Proud Boys channels, we used Telethon, a Python-based interface for the Telegram API. Data provided by the Telegram API include channel and post-level metadata, as well as the content of the posts themselves.

Channel metadata provided by the Telegram API include the unique identification number, title, creation date, current count of users, and various channel settings (e.g., usage configurations, administrator restrictions, and whether the channel is a bot), as well as the actual messages sent in the channel. Messages can be either original content posted to a channel or forwarded content from another channel. We did not collect user handles pertaining to the individuals who posted on these channels.

We started with a seed list of 10 Proud Boys-affiliated broadcast channels on Telegram identified by leading organizations working in this domain.[2] To grow the list of channels, we relied on mentions of other channels and the "forwarding" feature within Telegram, whereby content can be shared between channels. Each time we encountered a new mention of a channel or content forwarded from a channel that was not already on our list, we added the channel to our list, collected its data, and followed all of its channels. This snowball "crawling" approach resulted in a list of more than 2,900 unique channels that are closely affiliated with the Proud Boys.

This network, depicted in Figure 1, was created by using all forwardings from the 571 channels with full histories (self-citations were included). Forwardings include direct re-posts from other channels/groups (sometimes with their added comments) in message texts. It has 2,934 nodes. The node size represents eigenvector centrality, which measures a node's importance while giving consideration to the importance of its neighbors. The larger the node size is, the greater influence the node has.

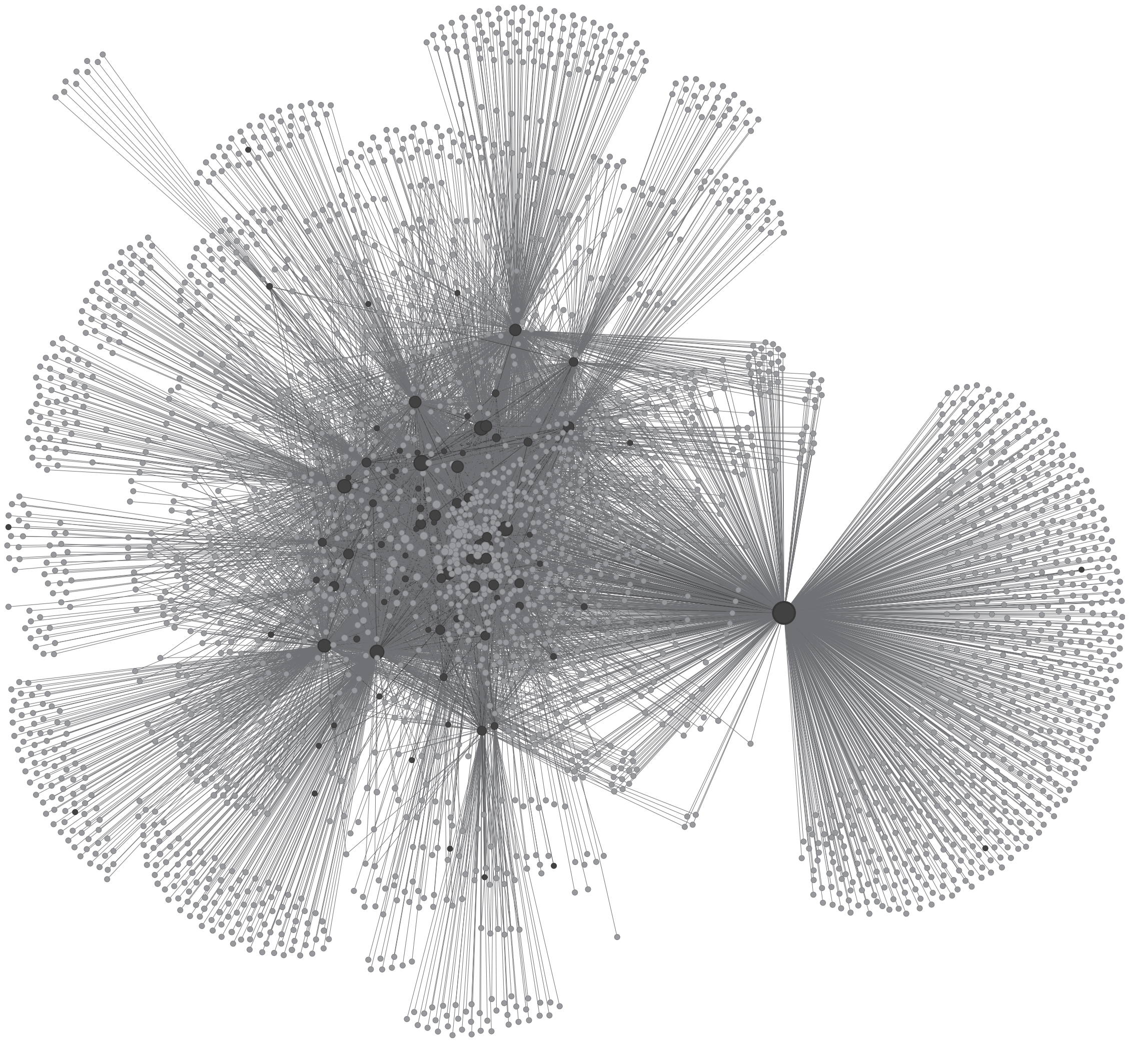
The dataset for this analysis comprises the text of 514,368 messages posted between January 1, 2020 and July 26, 2022 in 92 channels that explicitly identify as

[1] Please visit https://telegram.org/faq for more information.

[2] The list of channels that served as the starting point for our data collection was shared by Moonshot (https://moonshotteam.com).

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

FIGURE 1. Network Map of Proud Boys-Affiliated Telegram Channels



Proud Boys-affiliated in their username, title, or bio.[3] Although this is a relatively small portion of the full list of Proud Boys-aligned channels, these 92 channels form the core of the larger communication network on Telegram. To illustrate this, Figure 1 shows the message-forwarding network among all 2,934 channels. The nodes are colored by type: Black nodes represent explicitly Proud Boys-affiliated channels, and gray indicates other far-right related channels that promote Proud Boys-aligned content. The size of the node is based on how often messages were forwarded from that channel to another within the network. Almost all of the explicitly Proud Boys-affiliated channels are near the center of the network graph and are also comparatively large, meaning these channels serve as important hubs for sharing Proud Boys-related content with the larger far-right Telegram ecosystem.

Computational Approach to Collective Action Frame Classification

Most work focused on developing computational models for classifying extremist communication online has been shaped by a content moderation lens. That is, this work has been motivated largely by an interest in identifying and in turn responding to—for example, by removing, hiding, or down-ranking—harmful content and those who produce it, especially using automated means (Ahmed, Vidgen, and Hale 2022; Kiritchenko, Nejadgholi, and Fraser 2021). This has led to particular focus on concepts such as abusive language, toxicity, and hate speech (Bianchi et al. 2022; Davidson et al. 2017; Mercan et al. 2021; Schmidt and Wiegand 2019; Waseem and Hovy 2016). Social media platforms themselves tend to shape their policies around manifestly harmful communication (Meta 2022). In most instances, extremist communication is treated as a discrete speech act, with methods and interventions attending to individual posts.

[3] Please refer to the Supplementary material, Bailard et al. (2024), and Zhong et al. (2024) for more information about this network and its channels.

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

These automated approaches have grown in sophistication over time, especially as interdisciplinary work has helped generate models rooted in richer and more nuanced conceptual frameworks (Bianchi et al. 2022; Vidgen et al. 2021). This body of work has also helped to identify and address some of the most unambiguously worrisome content online. However, less attention has been paid to types of communication that may lead to harmful outcomes but are not themselves, *prima facie*, hateful or abusive.

As laid out in the aforementioned theoretical framework, it is likely that extremist groups' adherents are mobilized to action by a variety of frames that are part of relatively mundane, everyday discourse. And as such, if we are to better understand whether and how right-wing extremist groups' online communication is related to offline mobilization, we need computational models that allow us to capture and analyze these more subtle and complex communicative dynamics. With this in mind, we have fine-tuned a state-of-the-art natural language processing (NLP) model, relying on supervised machine learning techniques, to identify collective action frames at scale.

Frame Annotation

In order to identify collective action frames within the messages in our dataset, we fine-tuned a pretrained language model using a gold label dataset of 12,189 messages labeled by a team of five trained undergraduate and graduate students at George Washington University. Annotators applied labels on the basis of a detailed codebook developed by two of the study's authors. Pairs of students independently annotated posts in batches. The unit of analysis was the Telegram post, meaning that a single post could contain more than one frame, and each post was annotated with a binary "yes/no" for the presence or absence of a specific frame. Any disagreements between the paired annotators were identified, and the students met to resolve their disagreements. In order to prevent discrepancies from developing across annotators, the student pairs rotated with each batch, with one student designated in each rotation to attend the resolution meetings, observing and sharing any apparent discrepancies with the full team. The team then collectively agreed on a standard and clarified the codebook where needed.

The period in which these standards were being set and updated in the guidelines was treated as a training phase. Data from the training phase were re-annotated once the guidelines were settled. Once inter-annotator agreement was consistently high (above 0.7 for Gwet's A.C. and 80% for agreement, with averages of 0.77 and 86% across all five labels), the team began the full annotation phase. The procedures remained the same during this phase, with one student continuing to observe resolution meetings. We measured inter-annotator agreement for every batch of messages, with averages of 0.82 Gwet's A.C. score and 88% agreement across the full gold label dataset.

Computational Modeling and Analysis

Pretrained language models (Devlin et al. 2019; Radford et al. 2019) are now among the most prominent techniques in NLP. These models are built with deep neural networks that are pretrained on corpora made of billions of words in a language modeling task: they learn to predict the next words (or missing words) from pieces of text. After pretraining, these models can be adapted to solve many tasks, such as classification, in many different languages with great performance (Nozza, Bianchi, and Hovy 2020). Thus, we followed a common paradigm in current NLP research by making use of one such pretrained model—DeBERTa (He, Gao, and Chen 2021), which has shown excellent performance in text classification tasks—and fine-tuned the model for our specific classification task—that is, detecting and labeling collective action frames.

Preprocessing We applied only minor cleaning tasks to the data. We did not remove links from the text, but we split them to separate the words in the subdirectory part of the URL. For example, the link "www.shop.com/buy-our-stuff" was transformed to "www shop com buy our stuff." Minor cleaning is applied to both the data for tuning the model and the dataset for the final analysis.

Model DeBERTa is trained on a multi-label classification task—meaning that, given a message as an input, the model is trained to predict one or more labels that occur within the message, at the same time.

We trained the model on the collective action frames. In keeping with the standard approach in machine learning, we randomly divided the data into three sets: training (80% of the data), validation (10% of the data), and test (10% of the data). The training set was used to start training the model, while the validation set was used to select at which point, during the training, the model reached its best performance. Validation was run every 100 steps, with a total batch size of eight with a gradient accumulation of eight using a learning rate of 5e-5. The learning rate was identified using a grid search over the following set of values [1e-5, 5e-6, 8e-6, 9e-6, 5e-5].[4]

Finally, we checked the results using the test set to ensure the macro-F1 scores of the labels met the accepted thresholds.[5] As an additional check of these results, we randomly sampled 376 previously unseen posts that were then labeled by one of the same student annotators who labeled the training set, who was blind to the results of the computational analysis. A comparison of the labels predicted by the computational model to those assigned by the student annotator yielded percent agreement rates ranging from 85% to 97% for the five categories, with an average Gwet's AC score of 0.91.

[4] For more information about the computational methodology, please refer to the Supplementary material.

[5] Macro average F1 was equal to 0.80, while label-specific F1s for diagnostic, prognostic, motivational, othering, and injustice were 0.85, 0.85, 0.78, 0.76, and 0.78, respectively.

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

Statistical Analysis

To investigate how the Proud Boys deploy various collective action frames ahead of and in response to offline activities, we merged the manually and computationally labeled datasets of Telegram messages with a subset of ACLED's US Crisis Monitor data (ACLED 2019; Clionadh et al. 2010), consisting of 376 events in which members of the Proud Boys were identified as participating in between January 1, 2020, and July 26, 2022. In some cases, the Proud Boys organization is the primary or only organized group identified as participating in an event, in other cases they are one of the multiple groups.

The specific subcategories of Proud Boys-affiliated events identified in the ACLED data during this time period include arrests (7), attacks (2), changes to group/activity (1), peaceful protests (268), protests with (nonviolent) intervention (19), looting/property destruction (1), violent demonstrations (56), and events classified as "other" (22).[6] For the purpose of our analysis, we focus on two overarching categories, nonviolent protests and violent events. Violent events include those categorized as attacks, violent demonstrations, or looting/property destruction, whereas nonviolent protests include peaceful protests and protests that may have entailed some type of interaction with another group and/or authorities, but for which no known acts of violence or injuries occurred.[7]

We collapsed both this subset of ACLED data and our Telegram dataset into the same seven-day periods, and then merged these together into a single time-series dataset consisting of weekly measures of the number of nonviolent protests Proud Boys members participated in, the number of violent events that Proud Boys members participated in, the percentage of Telegram posts that included each type of collective action frame, and the number of posts (logged) containing each type of frame during a given week.[8] This data structure permits time series analyses of weekly data to investigate the temporal correlation between and uni- or bidirectional predictive power of the prevalence and proportion of specific types of collective action frames with Proud Boys members' participation in offline events.

Using Granger causality tests (Granger 1969), we examine whether specific collective action frames "Granger cause" offline events and vice versa. Granger causality tests are a staple in political communication investigations of intermedia agenda-setting effects. For example, studies have employed these models to analyze whether political blogs set the agenda for traditional news media outlets or vice versa (Meraz 2011), as well as the bidirectional agenda-setting relationships between traditional media outlets and both social media platforms (Groshek and Groshek 2013; Su, Hu, and LaiLee 2020) and aggregate trends in online searches (Ragas, Tran, and Martin 2014).

Despite the model's name, Granger causality tests do not test for causality in the theoretical sense—that is, they do not empirically substantiate that a specific factor *caused* another—nor do we seek to demonstrate causality in this research. Rather, Granger causality tests the *predictive power* or *incremental forecasting value* of one variable (and its lags) on another variable in the model. In other words, it assesses whether including a specific variable improves the model's ability to accurately predict subsequent levels of another variable.

We complement these findings with impulse response functions (IRFs), which trace the impact of a one-standard-deviation "shock" (i.e., impulse) in one variable on the subsequent behavior of other variables in the model and itself (Hamilton 1994; Lütkepohl 2005). Thus, IRFs also do not test for a causal relationship between the variables but, rather, calculate and map the interrelationships within and between variables over time.[9]

To generate the Granger causality and IRF results, we begin with vector autoregressive (VAR) models, which model a vector of variables dependent on their own lags as well as on the lags of the other variables included in the vector (Lütkepohl 2005). Before conducting the VAR tests, we examined the data to ensure they adhered to the assumptions integral to this type of analysis. First, we employed Akaike information criterion (AIC) and final prediction error (FPE) tests to identify the optimal lag lengths (Akaike 1988; Takeshi 1985), which indicated up to two weeks as the optimal lag length across the models.[10] After doing so, we conducted augmented Dickey–Fuller (ADF) tests, in which all of the variables exhibited stationarity at the optimal number of two lags (Dickey and Fuller 1979; Hamilton 1994). We next tested the results of the VAR models for stability (Hamilton 1994; Lütkepohl 1993), as well as the absence of autocorrelation of the residuals using Lagrange multiplier (LM) tests (Davidson and MacKinnon 1993; Johansen 1995). These tests

[6] For more information, please refer to the US Crisis Monitor codebook, FAQs, and dataset (see ACLED 2019; https://acleddata.com).

[7] We exclude the subcategory of arrests, changes to group/activity, and those classified by ACLED as "other" for the purpose of the present analysis, as they are not a straightforward conceptual fit for either of our primary categories of violent events and nonviolent protests.

[8] Due to being highly right-skewed, the frequency measures are log-transformed for the purpose of this analysis.

[9] Our results are reported in the form of orthogonalized IRFs (Lütkepohl 2010, 145).

[10] FPE and AIC tests indicate up to two weeks as the optimal lag length, with the exception of two VAR models: (1) nonviolent events and percent diagnostic, prognostic, and motivational variables and (2) violent events and percent injustice, prognostic, and motivational. For each, one week is indicated as the optimal lag. However, in both cases, LM tests of the VAR models at one lag show significant autocorrelation of the residuals, whereas, at two lags, the null of no autocorrelation cannot be rejected. The AIC scores are also lower for both at two lags compared to one lag. For these reasons, as well as consistency across the models, two lags are used for all analyses. However, we also ran these VAR models with one lag, and the results were commensurate. (Please refer to the primary analysis script document in the Dataverse for more details.)

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

confirm that the data satisfactorily meet the conditions for a VAR analysis at level for two lags.[11]

FINDINGS

Our analysis reveals a clear correlation between the percentage and frequency of posts to Proud Boys Telegram channels that include one or more of the collective action frames and the participation of members in offline events. Before discussing our results in more detail in the following subsections, however, it is useful to reiterate that these models do not test for a *causal* relationship in the theoretical sense. Rather, our objective is to test whether these variables are temporally correlated with one another, such that changes to one or more of the variables forecast or predict subsequent behavior or trends of the other variables. Thus, in the context of this longitudinal analysis, the term "predict" does not entail any claim of an empirical test of causality. This approach instead investigates how Proud Boys' online discussions relate to their offline activities by illuminating the types of discursive appeals they use while mobilizing ahead of and in the aftermath of offline events.

Overview

We consider both the sheer number (i.e., frequency) of messages and the percentage (i.e., proportion) of messages featuring a given frame per week, each of which captures a meaningful dimension of the Telegram information ecosystem and conversations happening between members of the Proud Boys. Whereas frequency captures the extent and magnitude of a particular type of conversation, the relative percentage of posts featuring a given frame captures the tenor and focus of the Proud Boys discourse at a given point in time.[12]

Beginning with an overview of the presence of these frames in Proud Boys conversations on Telegram, each of the collective action frames was a regular component of messages posted to these Telegram channels. Diagnostic frames are the most prevalent, appearing in an average of 34% of the weekly posts to the 92 channels included in our analysis, with averages of 13% of the weekly messages including an injustice frame and 11% including an othering frame. The next most common type of frame is prognostic, which appeared in an average of 12% of the weekly posts, followed by motivational frames at 9%. Moving to frequency, for the 92 channels comprising our core network of Proud Boys-affiliated channels, we find an average of 3,839 messages posted per week—ranging from a minimum of 356 messages to a maximum of 49,228 total messages in a specific week—with weekly averages of 1,185 messages featuring diagnostic frames, 423 featuring prognostic frames, and 293 featuring motivational frames.

Nonviolent Protests

The results of our analysis of nonviolent protests fail to substantiate the mobilizing capacity of any of the three collective action frames hypothesized by the social movement literature—neither the proportion nor frequency of any of these frames predicts participation in an impending nonviolent protest. However, nonviolent protests strongly predict an increase in the percentage of messages that feature a motivational frame over the following weeks. (Please see Table 1 and Figure 2 for these results.) This raises the possibility that offline nonviolent protests may function to provide fodder for or a focal point that shapes subsequent online conversations between members of the Proud Boys, the potential implications of which are discussed in the following sections.

Violent Events

In contrast to nonviolent protests, violent events are strongly predicted by an increase in the percentage of posts that include either a diagnostic or a motivational frame, as well as an increase in the number of messages that include a motivational frame. (Please see Table 2 and Figure 3.) These results substantiate Hypotheses 1 and 3—messages that lament or assign blame for some problem perceived to afflict the Proud Boys, or that positively prime in-group identity by emphasizing pride, solidarity, or morale, each increase the likelihood that Proud Boys members will participate in a violent event in the coming weeks. Turning to the diagnostic subframes, an increase in the percentage of messages that contain either an injustice or othering frame predicts an increased likelihood of a violent event in the coming weeks ($p \leq 0.07$ and $p \leq 0.09$, respectively)—in line with Hypotheses 1A and 1B.[13] (Please see Figure 4.)

[11] For the full results of all VAR tests for the Granger causality results (Tables 1–3) and IRFs (Figures 2–5) reported here, please see the Supplementary material.

[12] 14 of these channels self-identify as being outside of the United States (e.g., Australia, Canada, etc.). Since we allowed the shape of the network itself to determine which channels constitute the network's core, we opted to include these non-US channels in the primary analysis. Additionally, the content and conversations circulating through these channels are not geographically constrained, nor does our theoretical framework require it to be. Nevertheless, we conducted an additional analysis excluding non-US channels, and the primary findings remained commensurate with the main analysis. (Please see the Supplementary material for these results.)

However, unlike the primary analysis reported here, nonviolent protests predict an increase in the number of diagnostic and prognostic messages ($p < 0.01$) in the analysis excluding non-US channels. Although it is possible and likely that non-US citizens interact on the remaining channels, the finding that nonviolent protests predict an increase in the frequency of these types of messages in a seemingly more US-centric network merits additional analysis in future investigations.

[13] Please see Supplementary Table A3 for the results of the tests of the injustice and othering frames.

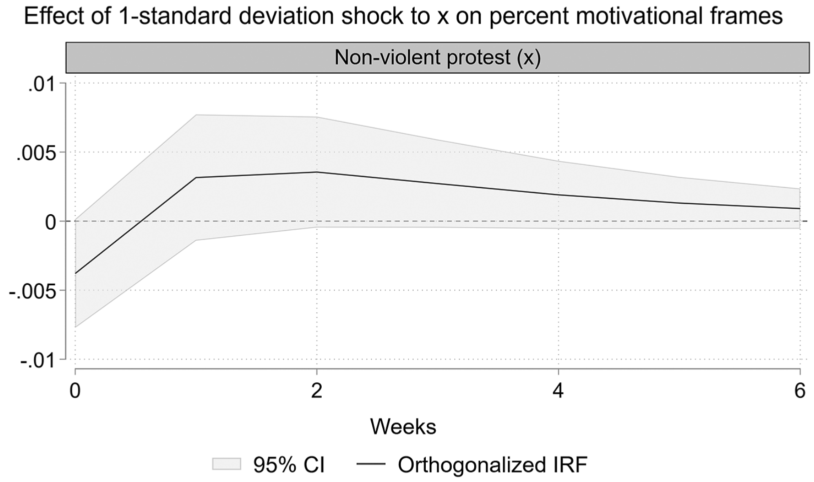
https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

TABLE 1. Results of Granger Causality Tests of Nonviolent Events and Collective Action Frames

Nonviolent Protests							
Percent of posts containing frame				*Number (logged) of posts containing frame*			
Granger cause ->		Chi^2	$Prob>chi^2$	Granger cause ->		Chi^2	$Prob>chi^2$
Diagnostic	Nonviolent protests	0.17	0.92	Diagnostic	Nonviolent protests	0.91	0.64
Prognostic	Nonviolent protests	0.64	0.73	Prognostic	Nonviolent protests	1.41	0.5
Motivational	Nonviolent protests	2.19	0.33	Motivational	Nonviolent protests	1.75	0.42
Nonviolent protests	Diagnostic	1.49	0.48	Nonviolent protests	Diagnostic	0.65	0.72
Nonviolent protests	Prognostic	0.54	0.76	Nonviolent protests	Prognostic	1.33	0.51
Nonviolent protests	Motivational	7.1	0.03**	Nonviolent protests	Motivational	1.1	0.58

Note: Significance levels indicated as *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

FIGURE 2. IRF Plot of Nonviolent Events on Percentage of Motivational Frames



Note: Please see Supplementary Tables A6–A9 for full specifications of the IRF results depicted in Figures 2–5.

Regional Analysis: A Robustness Check

To further investigate the predictive power of diagnostic and motivational frames for violent offline events, we shift the unit of analysis to a time series cross-sectional test of this correlation at the state level. Ideally, a geographically focused approach would entail a spatial analysis of geolocated Telegram data. However, correlating social media data with geographic outcomes remain prohibitively problematic, primarily due to the lack of data containing geolocation information and because self-reported user locations are often misleading (Hecht et al. 2011). Public channel data exposed by the Telegram API include metadata such as the unique identification number, title, and creation date, but not geographic location. Telegram users can opt in to provide their geolocation; however, very few do so. Moreover, because users self-select to provide their geolocation, this subset of data is biased in a manner that would be problematic for an analysis on this topic. Taken together, the sparse and unreliable nature of geolocated Telegram data renders it difficult to map online communication to offline events in a valid manner.

Additionally, our network analysis of the Proud Boys-affiliated channels revealed that this network is highly connected, with a mean undirected average clustering coefficient of 0.692 (Hansen et al. 2020).[14] On average, about 34% of these Proud Boys-affiliated channels' content was forwarded from other channels, including 11.4% directly from other explicitly Proud Boys-affiliated channels, highlighting this group's highly networked nature on Telegram. Thus, online text messages of Proud Boys channels were not independent but rather strongly influenced by other channels—further demonstrating the limitations associated

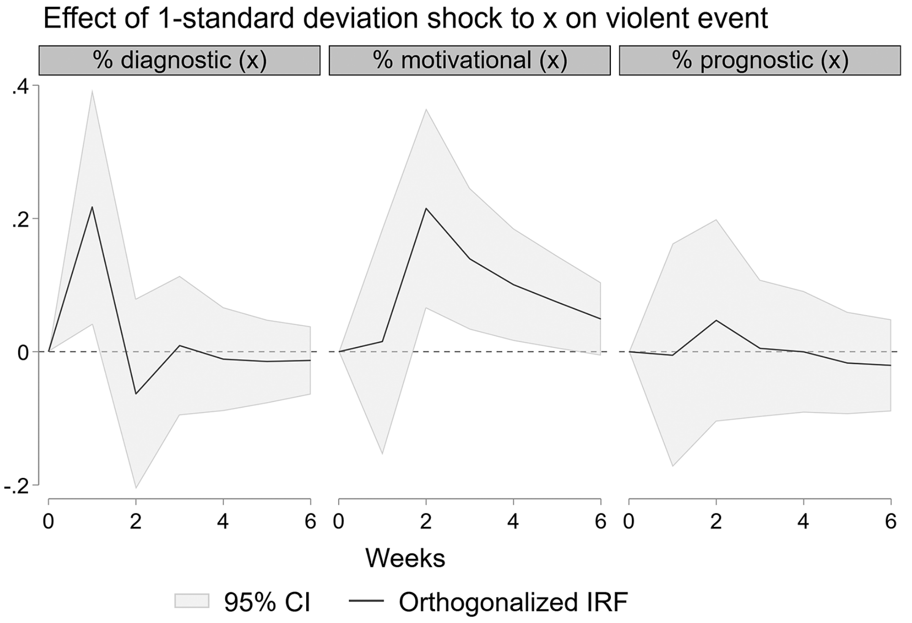
[14] A large coefficient (i.e., close to 1) indicates that a node is highly interconnected with its neighbors.

TABLE 2. Results of Granger Causality Tests of Violent Events and Collective Action Frames

Violent events							
Percent of posts containing frame				*Number (logged) of posts containing frame*			
Granger cause ->		Chi^2	Prob>chi^2	Granger cause ->		Chi^2	Prob>chi^2
Diagnostic	Violent events	6.68	0.04**	Diagnostic	Violent events	2.29	0.32
Prognostic	Violent events	1.44	0.49	Prognostic	Violent events	0.7	0.7
Motivational	Violent events	8.37	0.02**	Motivational	Violent events	6.43	0.04**
Violent events	Diagnostic	2.55	0.28	Violent events	Diagnostic	0.2	0.9
Violent events	Prognostic	3.1	0.21	Violent events	Prognostic	0.24	0.89
Violent events	Motivational	0.05	0.97	Violent events	Motivational	1.36	0.51

Note: Significance levels indicated as *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

FIGURE 3. IRF Plots of Violent Events and Percentage of Diagnostic, Motivational, and Prognostic Frames



with mapping interdependent online communication with localized offline events.

On the theoretical side, rather than explicit calls-to-action (i.e., prognostic frames) being predictive of offline events in this context, it is messages that lament or assign blame for some problem perceived to afflict the Proud Boys (i.e., diagnostic frames) or that positively prime in-group identity (i.e., motivational frames) that predict Proud Boys members' offline participation in a violent event. Thus, it is likely that the conversations and messages that motivate individuals to participate in offline activities are not geographically constrained—messages lauding the participation of Proud Boys in a protest in Seattle, or which bemoan some offense or injustice suffered by Proud Boys members in Idaho, could feasibly inspire offline activity by Proud Boys members in Minnesota.

Being mindful of these methodological and theoretical limitations, an analysis of this relationship at the regional level would nevertheless provide a useful robustness check of our main findings. To this end, we conducted a time series cross-sectional analysis of the 53 channels that explicitly self-identify with a particular region (e.g., "Houston Proud Boys" or "Indiana Proud Boys"), employing a first-difference model to test the correlation between the specific types of messages on these regional channels and participation in violent activities by Proud Boys in those respective states.[15] The results of this additional analysis further

[15] There are several features of the regional time series cross-sectional dataset that favor a first-difference model over fixed effects. The data are highly unbalanced as a result of channels being created

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

TABLE 3. First-Difference Test of the Effect of the Number of Messages (x) on the Occurrence of Violent Events (y) at the Regional Level (Clustered Standard Errors)

	Coefficient (S.E.)	Z-score	*p* value
Number of diagnostic frames	0.0004 (0.0005)	0.74	0.46
Number of prognostic frames	–0.0006 (0.001)	–0.58	0.56
Number of motivational frames	0.002 (0.001)	2.14	0.03**
Constant	–0.0003 (0.0006)		

Note: Total observations = 3,099; number of groups = 53; range of observations/group = 11–133; $r^2 = 0.003$. Significance levels indicated as $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

substantiate the correlation between motivational frames and violent offline events. In this case, the number of messages featuring a motivational frame shared on a regional channel is correlated with Proud Boys members' participation in violent offline events in that state ($p < 0.05$). However, the correlation between diagnostic frames and violent offline events is nonsignificant. (Please see Table 3 for results.)

An Online Messaging–Offline Action Cycle

Considering the analyses of nonviolent protests and violent events in tandem with one another illuminates a potential reciprocal relationship shared by online communication and offline events. Whereas Proud Boys members' participation in nonviolent protests increases the percentage of motivational frames in their Telegram posts over the following weeks, this boost in the percentage and number of motivational frames, in turn, increases the likelihood that Proud Boys members will engage in an upcoming violent event. This raises the question of whether nonviolent protests may help lay the groundwork for later violent activities, in part, by shaping the tenor and focus of online conversations between group members. Nonviolent protests, for example, could provide fodder for online rallying cries and appeals to members' sense of group pride, duty, or morale in the days and weeks following a protest. Additionally, these protests may function as a tangible show of force, increasing members' sense of their strength in numbers and solidarity. As posited by collective action theory, these types of discursive appeals can recalibrate members' cost–benefit calculus, such that participation in risky violent activities becomes more likely to be deemed as worthwhile, which we discuss further in the following section.[16]

In a preliminary investigation of this potential reciprocal dynamic, superimposing the IRFs for nonviolent protests with percent motivational frames and violent events suggests a four-week timeline for the temporal relationships shared by these factors. (Please see Figure 6.) Whereas the effect of a one-deviation shock to nonviolent protests on percent motivational frames peaks at two weeks after the shock, the predicted effect of a one-deviation shock to motivational frames on violent events peaks two weeks after that. This is complemented by an IRF test of the direct temporal relationship shared by nonviolent protests and violent events, which reveals that nonviolent events predict the occurrence of a violent event four weeks later. (Please see Figure 5.)

Utilizing a structural equation model, we tested whether the increase in the percentage of motivational frames following nonviolent protests mediates their predictive power for subsequent violent events. (Please see Table 4.) A modified Baron and Kenney test (Iacobucci, Saldanha, and Deng 2007) finds that the effect of nonviolent protests on violent events is partially mediated by the increase in the percentage of messages that contain a motivational frame at the 0.10 significance level (X - > M: B = 0.002 and p = 0.10; M - > Y: B = 5.471 and p = 0.06; Sobel's z-test: B = 0.01, p = 0.21), with an estimated 8% of the direct effect of nonviolent events on violent events mediated by the increase in the percentage of motivational messages. Although it should be stipulated that mediation analyses are limited by a number of caveats (Agler and De Boeck 2017)—and, thus, these results should be considered with a grain of salt—this provides a starting point to investigate whether nonviolent protests are leveraged in online communication to make discursive appeals that increase group members' willingness to participate in subsequent violent events.

Theoretical Implications

We did not undertake the aforementioned analysis with theoretically derived expectations in mind regarding the ways in which offline activities might impact online communication. Though scholars in the collective action framing tradition view framing as an act of meaning-making, this body of work has given more attention to the ways in which frames spur action. Yet our longitudinal analysis suggests a cyclical pattern between offline

and deleted at different points during the time span of the analysis. Additionally, modified Wald and LM tests reveal significant groupwise heteroskedasticity of the residuals and serial correlation of the errors. Finally, the data include a moderate number of units and a larger number of time periods (i.e., T > N) than is typical for analyses of panel data. In each of these regards, first-difference models are better-suited to an analysis of a dataset with these properties compared to fixed effects models (Woolridge 2010; 2015).

[16] It is likely that some of these violent events were not originally planned nor intended to be violent, but nevertheless resulted in some form of violence. Regardless, it is feasible that an increased focus on grievances and/or motivational appeals in online communication fosters a mindset that increases members' willingness to commit violent acts, whether or not they were explicitly preplanned.

FIGURE 4. IRF Plots of Violent Events and Percentage of Injustice and Othering Frames

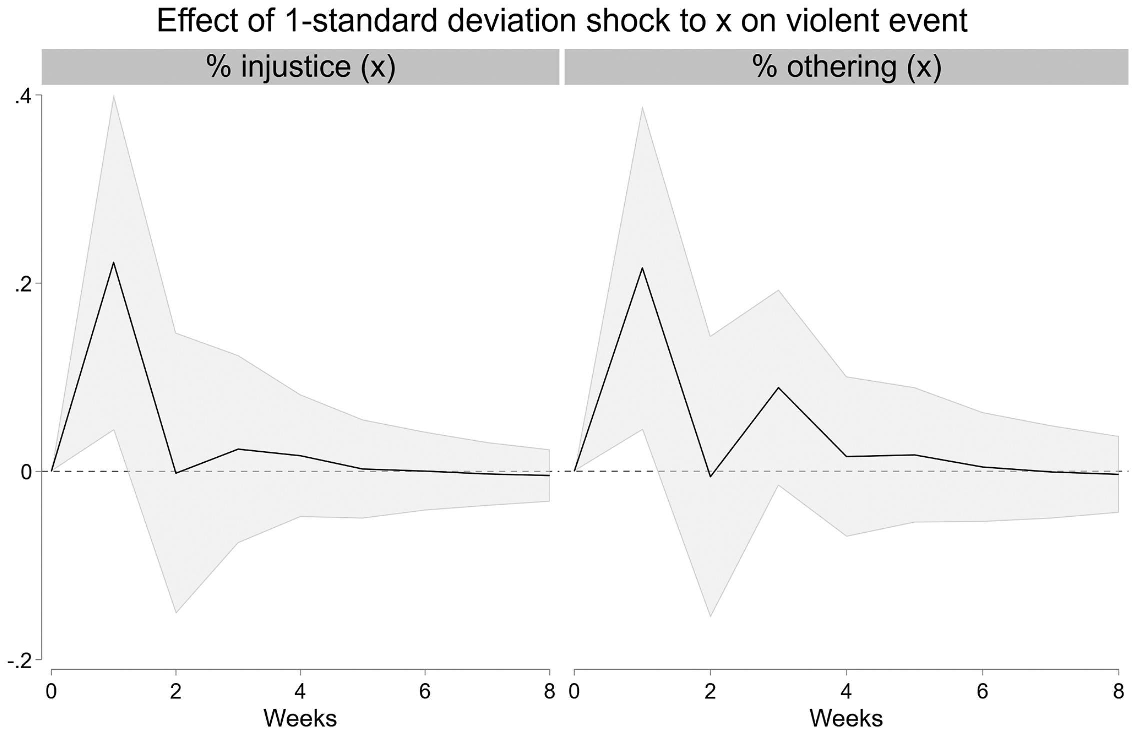
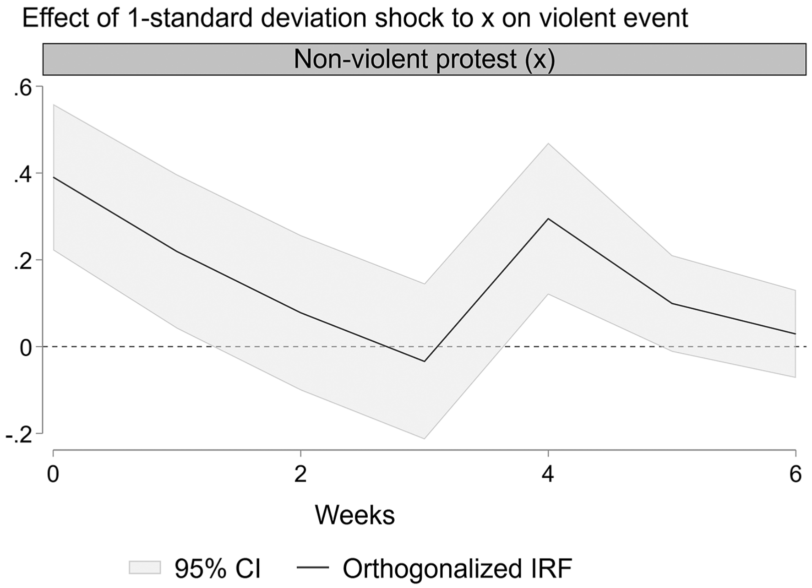


FIGURE 5. IRF Plot of Nonviolent Events and Violent Events



activities and certain online speech, which we believe merits further exploration in future research.

These findings also have potential theoretical implications for the social movement literature, as well as our understanding of the link between online communication and offline participation. In this regard, we suggest supplementing collective action framing theory with insights drawn from social and political psychology on "moralizing" and "moral convergence." This body of work finds that attaching moral reasoning

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

FIGURE 6. IRFs of Percent Motivational Frames, Nonviolent Events, and Violent Events

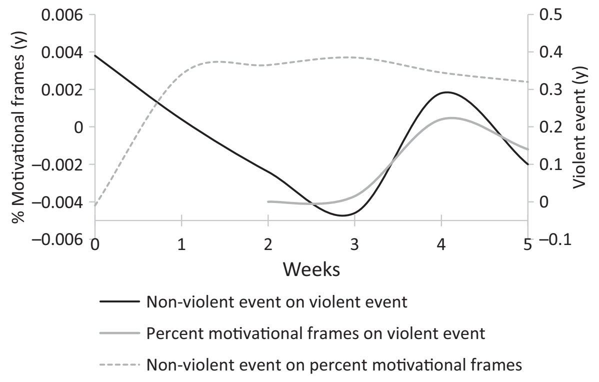


TABLE 4. Results of the Structural Equation Model

Paths: Explanatory - > Outcome	Coefficient (S.E.)	Z-score	*p*-value
Nonviolent event (t_{-4}) - > % motivational messages (t_{-2})	0.002 (0.001)	1.65	0.1*
% motivational messages (t_{-2}) - > violent event (t_0)	5.47 (2.87)	1.91	0.06*
Nonviolent event (t_{-4}) - > violent event (t_0)	0.14 (0.04)	3.2	0.001***

Note: *N* = 130; coefficient of determination = 0.09; significance levels indicated as **p* < 0.10, ***p* < 0.05, ****p* < 0.01.

to—that is, moralizing—an issue or cause is associated with the following:

> (a) greater preferred social and physical distance from attitudinally dissimilar others, (b) intolerance of attitudinally dissimilar others in both intimate (e.g., friend) and distant relationships…, (c) lower levels of good will and cooperativeness in attitudinally heterogeneous groups, and (d) a greater inability to generate procedural solutions to resolve disagreements (Skitka, Bauman, and Sargis 2005, 895).

Such social distance and unwillingness to cooperate, in turn, make risky behavior more palatable (Mooijman et al. 2018). For extremist groups, who, by definition, operate outside of the mainstream, offline mobilization is always at least somewhat risky, and thus, attaching moral reasoning to their activities is essential. Indeed, as our data suggest, these moral messages—in the form of diagnostic frames—are nearly omnipresent in online discussions between group members. We argue that these types of messages are key components of how a group conceptualizes its position and plight in an unjust world—moralizing prerequisites for mobilization. However, for mobilization to tip into violence, simple moralizing may not be enough. Support for violence is more likely when "moral convergence" has also taken place (Mooijman et al. 2018)—that is, when members believe that others within the group share these moral values. We argue that motivational frames, which positively prime in-group identity by emphasizing a group's shared values, pride, strength and/or solidarity, constitute one such avenue for moral convergence.

One of the best opportunities for members of extremist groups to develop a sense of moral convergence, we suggest, is via offline interaction—such as at protest events, where members see firsthand that others think and feel as they do. They bond over shared ritual and risk. However, these same events are also subsequently recounted and celebrated online, which increases the groups' sense of solidarity and camaraderie more broadly. Following an offline event, then, feelings of solidarity precipitated by offline activity translate to increased expressions of that solidarity online in the form of motivational frames, which enhances the sense of moral convergence for members more broadly. And with these perceptions heightened for more adherents, the willingness to engage in violence may itself grow stronger—resulting in the *online messaging–offline action cycle* represented in our empirical data.

CONCLUSION

The results of our empirical analysis support key insights drawn from social movement and collective action framing theory—both diagnostic and motivational messages prove predictive of Proud Boys' participation in violent offline events, substantiating the capacity for these collective action frames to help mobilize group members. However, our results also point to the need for further theoretical development and testing. Not only are prognostic frames uncorrelated with the Proud Boys' offline activities, but the relationships between diagnostic and motivational frames, on the one hand, and violent and nonviolent events, on the other, prove more complex than hypothesized. Combining collective action framing theory with insights from research on moralizing and moral convergence, we have offered an enhanced theoretical framework that attends to the cyclical nature of the observed relationship between online messaging and offline action. However, this inductively derived framework will require much more testing in future research into different extremist groups, platforms, and time frames.

Our findings also have important implications from a practical, content moderation perspective. While previous approaches to large-scale content detection and moderation have primarily focused on uncivil, hateful, and other toxic content tied to right-wing extremism (Ahmed, Vidgen, and Hale 2022; Bianchi et al. 2022), by shifting attention to the Proud Boys' use of collective action frames over time, we have shown how social media can be harnessed by groups to cultivate a worldview and sense of in-group identity and solidarity that lays the groundwork for mobilizing its members to participate in offline activities. While these sorts of messages are often angry and contentious, they are not always explicitly hateful.

That prognostic frames, which most often take the form of explicit calls-to-action, are not predictive of offline events in our analysis also has practical implications, particularly because these types of messages receive considerable attention from social media platforms. From a technical perspective, such calls-to-action are among the easiest to automatically detect, and platforms are particularly likely to moderate posts that call for, especially violent, offline action. However, it is also the case that this type of communication is more likely to occur via private or even encrypted channels, which are typically outside the domain of content moderation efforts. To be clear, we do not contend that platforms should allow such posts to remain unmoderated. However, our findings reveal the limitations of approaches that focus relatively narrowly on specific posts and their explicit content.

To the degree that other right-wing extremist groups share similar worldviews and grievances, it is likely that the dynamics identified in this study could generalize to similar groups on Telegram, both within and outside of the United States. However, the cost–benefit calculus of engaging in violence is likely to be contingent on additional group resources as well as the national context—suggesting that, even if these underlying dynamics function similarly, the threshold point at which the potential benefits of violent activity exceed its risks will vary.

Regarding the degree to which our findings may generalize to other platforms, it is useful to consider several factors, such as platform features, moderation policies, user base, and the overall focus of the platform. For example, relative to platforms that require verified identities, a platform that allows anonymity may lead to more aggressive or controversial behavior (Suler 2004). Additionally, platforms with strict moderation policies may lead to more covert or coded language usage, while those with minimal moderation may encourage more overt and explicit extremist content (Bhat and Klein 2020). Finally, the composition of the user base and homophilic patterns of a platform's network structures may also moderate the content of online discussions, particularly in terms of the formation of echo chambers and the nature of the information shared within (Cinelli et al. 2021). Thus, although right-wing extremist activity may exhibit some similarities across different platforms, unique platform characteristics will influence the way extremist group members communicate, organize, and engage with one another.

While right-wing extremist groups do use social media platforms to plan logistics for various actions and events, they also mobilize through other, more subtle forms of communication. If we are to meaningfully enhance our understanding of the relationship between online communication by extremist groups and their offline behavior, we must take a more nuanced view of these communicative dynamics and how they develop over time. Using state-of-the-art computational techniques to capture the theoretically rich and nuanced concept of collective action framing—ultimately uncovering the complex, cyclical nature of these dynamics—we have shown how such work might progress.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0003055423001478.

DATA AVAILABILITY STATEMENT

Documentation, datasets, and all analysis scripts used in the statistical analyses are openly available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/OAOJQZ. The raw dataset containing the text of the Telegram posts, which was used in the hand-labeling and computational modeling procedures, will not be made publicly available. Although this dataset includes only text messages from public channels and no user's personally identifiable information was collected by the authors, it still may be possible to reidentify users. To protect privacy and comply with the EU's General Data Protection Regulation standards, these raw data will only be shared for limited research purposes and require a signed data-sharing agreement. Readers interested in requesting access to the raw data should email: datarequest.apsrstudy@gmail.com.

ACKNOWLEDGEMENTS

We first thank our incredible team of research assistants: Autumn Archuleta, Carolyne Im, Elizabeth Irwin, Olivia Legault, Max Zweig, and others. We also thank the Institute for Data, Democracy & Politics and the John S. and James L. Knight Foundation for their support of our work. Thanks also to the editors and reviewers for their valuable feedback, as well as to those individuals at the Bridging Divides Initiative, the Holocaust Museum, Moonshot, and ADL who shared their time to chat with us during the early brainstorming phase of this project. All mistakes are our own.

FUNDING STATEMENT

This research is funded by the John S. and James L. Knight Foundation through a grant to the Institute for Data, Democracy & Politics at The George Washington University.

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors affirm this research did not involve human subjects.

REFERENCES

ACLED. 2019. "Armed Conflict Location and Event Data Project (ACLED) Codebook."

Adams, Josh, and Vincent J. Roscigno. 2005. "White Supremacists, Oppositional Culture and the World Wide Web." *Social Forces* 84 (2): 759–78.

Agler, Robert, and Paul De Boeck. 2017. "On the Interpretation and Use of Mediation: Multiple Perspectives on Mediation Analysis." *Frontiers in Psychology* 8: Article 1984.

Ahmed, Zo, Bertie Vidgen, and Scott A. Hale. 2022. "Tackling Racial Bias in Automated Online Hate Detection: Towards Fair and Accurate Classification of Hateful Online Users Using Geometric Deep Learning." *EPJ Data Science* 11 (8). https://doi.org/10.1140/epjds/s13688-022-00319-9.

Akaike, Hirotugu. 1988. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Selected Papers of Hirotugu Akaike*, eds. Emanuel Parzen, Kunio Tanabe and Genshiro Kitagawa, 199–213. New York: Springer.

Al-Rawi, Ahmed. 2021. "Telegramming Hate: Far-Right Themes on Dark Social Media." *Canadian Journal of Communication* 46 (4): 821–51.

Anti-Defamation League. 2018. "Backgrounder: Proud Boys." January 23. https://www.adl.org/resources/backgrounders/proud-boys-0.

Bailard, Catie Snow, Rebekah Tromble, Wei Zhong, Federico Bianchi, Pedram Hosseini, and David Broniatowski. 2024. "Replication Data for: "Keep Your Heads Held High Boys!": Examining the Relationship between the Proud Boys' Online Discourse and Offline Activities." Harvard Dataverse. Dataset. https://doi.org/10.7910/DVN/OAOJQZ.

Benford, Robert D., and Scott A. Hunt. 1992. "Dramaturgy and Social Movements: The Social Construction and Communication of Power." *Sociological Inquiry* 62 (1): 36–55.

Benford, Robert D., and David A. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual Review of Sociology* 26: 611–39.

Berbrier, Mitch. 1998. "'Half the Battle': Cultural Resonance, Framing Processes, and Ethnic Affectations in Contemporary White Separatist Rhetoric." *Social Problems* 45 (4): 431–50.

Berbrier, Mitch. 2000. "The Victim Ideology of White Supremacists and White Separatists in the United States." *Sociological Focus* 33 (2): 175–91.

Bhat, Prashanth, and Ofra Klein. 2020. "Covert Hate Speech: White Nationalists and Dog Whistle Communication on Twitter." In *Twitter, the Public Sphere, and the Chaos of Online Deliberation*, eds. Gwen Bouvier and Judith E. Rosenbaum, 151–72. Cham, Switzerland: Palgrave Macmillan.

Bianchi, Federico, Stefanie Anja Hills, Patricia Rossini, Dirk Hovy, Rebekah Tromble, and Nava Tintarev. 2022. "It's Not Just Hate": A Multi-Dimensional Perspective on Detecting Harmful Speech Online." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 8093–99. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics.

Bos, Linda, Christian Schemer, Nicoleta Corbu, Michael Hameleers, Ioannis Andreadis, Anne Schulz, Desirée Schmuck, et al. 2020. "The Effects of Populism as a Social Identity Frame on Persuasion and Mobilisation: Evidence from a 15-Country Experiment." *European Journal of Political Research* 59 (1): 3–24.

Boulianne, Shelley. 2015. "Social Media Use and Participation: A Meta-Analysis of Current Research." *Information, Communication & Society* 18 (5): 524–38.

Bubolz, Bryan F., and Pete Simi. 2019. "The Problem of Overgeneralization: The Case of Mental Health Problems and U.S. Violent White Supremacists." *American Behavioral Scientist* https://doi.org/10.1177/0002764219831746.

Busby, Ethan C., Joshua R. Gubler, and Kirk A. Hawkins. 2019. "Framing and Blame Attribution in Populist Rhetoric." *The Journal of Politics* 81 (2): 616–30.

Chong, Dennis, and James N. Druckman. 2007. "Framing Theory." *Annual Review of Political Science* 10: 103–26.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. "The Echo Chamber Effect on Social Media." *Proceedings of the National Academy of Sciences* 118 (9): e2023301118.

Clionadh, Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED-Armed Conflict Location and Event Data." *Journal of Peace Research* 47 (5): 651–60.

Davidson, Russell, and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Davidson, Thomas, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. "Automated Hate Speech Detection and the Problem of Offensive Language." *Proceedings of the International AAAI Conference on Web and Social Media* 11 (1): 512–5.

DeCook, Julia R. 2018. "Memes and Symbolic Violence:# Proudboys and the Use of Memes for Propaganda and the Construction of Collective Identity." *Learning, Media and Technology* 43 (4): 485–504.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 4171–86. Minneapolis, MN. Association for Computational Linguistics.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74 (366a): 427–31.

Dickson, Ej. 2021. "Proud Boys Channels are Exploding on Telegram." *Rolling Stone*, January 14. https://www.rollingstone.com/culture/culture-news/proud-boys-telegram-far-right-extremists-1114201/.

Gallacher, John D., Marc W. Heerdink, and Miles Hewstone. 2021. "Online Engagement between Opposing Political Protest Groups Via Social Media Is Linked to Physical Violence of Offline Encounters." *Social Media + Society* 7 (1): 2056305120984445.

Granger, Clive W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica: Journal of the Econometric Society* 37 (3): 424–38.

Groshek, Jacob, and Megan Clough Groshek. 2013. "Agenda Trending: Reciprocity and the Predictive Capacity of Social Network Sites in Intermedia Agenda Setting across Issues Over Time." Working Paper. https://ssrn.com/abstract=2199144.

Goffman, Erving. 1974. *Frame Analysis: An Essay on the Organization of Experience*. Boston, MA: Northeastern University Press.

Goh, Debbie, and Natalie Pang. 2016. "Protesting the Singapore Government: The Role of Collective Action Frames in Social Media Mobilization." *Telematics and Informatics* 33 (2): 525–33.

Greene, Viveca S. 2019. "'Deplorable' Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies." *Studies in American Humor* 5 (1): 31–69.

Guggenheim, Lauren, S. Mo Jang, Soo Young Bae, and W. Russell Neuman. 2015. "The Dynamics of Issue Frame Competition in Traditional and Social Media." *The Annals of the American Academy of Political and Social Science* 659 (1): 207–24.

Hamilton, James Douglas. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hansen, Derek L., Ben Shneiderman, Marc A. Smith, and Itai Himelboim. 2020. "Social Network Analysis: Measuring, Mapping, and Modeling Collections of Connections." *Analyzing Social Media Networks with NodeXL*: 31–51.

Hartzell, Stephanie L. 2020. "Whiteness Feels Good Here: Interrogating White Nationalist Rhetoric on Stormfront." *Communication and Critical/Cultural Studies* 17 (2): 129–48.

He, Pengcheng, Jianfeng Gao, and Weizhu Chen. 2021. "DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing." Working Paper.

Hecht, Brent, Lichan Hong, Bongwon Suh, and Ed H. Chi. 2011. "Tweets from Justin Bieber's Heart: The Dynamics of the Location Field in User Profiles." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 237–46. New York: Association for Computing Machinery.

Hunt, Scott A., Robert D. Benford, and David Snow. 1994. "Identity Fields: Framing Processes and the Social Construction of Movement Identities." In *New Social Movements: From Ideology* *to Identity*, eds. Enrique Laraña, Hank Johnston, and Joseph R. Gusfield, 185–208. Philadelphia, PA: Temple University Press

Iacobucci, Dawn, Neela Saldanha, and Xiaoyan Deng. 2007. "A Meditation on Mediation: Evidence that Structural Equations Models Perform Better than Regressions." *Journal of Consumer Psychology* 17 (2): 139–53.

Jackson, Sarah J., and Brooke Foucault Welles. 2015. "Hijacking #myNYPD: Social Media Dissent and Networked Counterpublics." *Journal of Communication* 65 (6): 932–52.

Johansen, Søren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Kawakami, Kerry, and Kenneth L. Dion. 1995. "Social Identity and Affect as Determinants of Collective Action: Toward an Integration of Relative Deprivation and Social Identity Theories." *Theory & Psychology* 5 (4): 551–77.

Khazraee, Emad, and Alison N. Novak. 2018. "Digitally Mediated Protest: Social Media Affordances for Collective Identity Construction." *Social Media + Society* 4 (1): 2056305118765740.

Kiritchenko, Svetlana, Isar Nejadgholi, and Kathleen C. Fraser. 2021. "Confronting Abusive Language Online: A Survey from the Ethical and Human Rights Perspective." *Journal of Artificial Intelligence Research* 71: 431–78.

Kreiss, Daniel, Regina G. Lawrence, and Shannon C. McGregor. 2020. "Political Identity Ownership: Symbolic Contests to Represent Members of the Public." *Social Media + Society* 6 (2): 2056305120926495.

Linton, Caroline. 2018. "Twitter Suspends Proud Boys, Gavin McInnes Accounts Ahead of the Unite the Right Rally." *CBS News,* August 10. https://www.cbsnews.com/news/proud-boys-gavin-mcinnes-twitter-suspension-today-unite-the-right-2018-08-10/.

Lütkepohl, Helmut. 1993. *Introduction to Multiple Time Series Analysis*. 2nd ed. Berlin, Germany: Springer.

Lütkepohl, Helmut. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer.

Lütkepohl, Helmut. 2010. "Impulse Response Function." In *Macroeconometrics and Time Series Analysis*, 145–50. London: Palgrave Macmillan

Mackie, Diane M., Thierry Devos, and Eliot R. Smith. 2000. "Intergroup Emotions: Explaining Offensive Action Tendencies in an Intergroup Context." *Journal of Personality and Social Psychology* 79 (4): 602.

Makki, Taj W., Julia R. DeCook, Travis Kadylak, and Olivia JuYoung Lee. 2018. "The Social Value of Snapchat: An Exploration of Affiliation Motivation, the Technology Acceptance Model, and Relational Maintenance in Snapchat Use." *International Journal of Human–Computer Interaction* 34 (5): 410–20.

Meraz, Sharon. 2011. "Using Time Series Analysis to Measure Intermedia Agenda-Setting Influence in Traditional Media and Political Blog Networks." *Journalism & Mass Communication Quarterly* 88 (1): 176–94.

Mercan, Vildan, Akhtar Jamil, Alaa Ali Hameed, Irfan Ahmed Magsi, Sibghatullah Bazai, and Syed Attique Shah. 2021. "Hate Speech and Offensive Language Detection from Social Media." In *International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 1–5. Quetta, Pakistan. doi: 10.1109/ICECube53880.2021.9628255.

Meta. 2022. "Violence & Incitement." https://transparency.fb.com/policies/community-standards/violence-incitement/ (accessed July 4, 2022).

McBain, Sophie. 2020. "The Rise of the Proud Boys in the US." *New Statesman*, October 7.

Mooijman, Marlon, Joe Hoover, Ying Lin, Heng Ji, and Morteza Dehghani. 2018. "Moralization in Social Networks and the Emergence of Violence during Protests." *Nature Human Behaviour* 2 (6): 389–96.

Müller, Karsten, and Carlo Schwarz. 2023. "From Hashtag to Hate Crime: Twitter and Antiminority Sentiment." *American Economic Journal: Applied Economics* 15 (3): 270–312.

Müller, Karsten, and Carlo Schwarz. 2021. "Fanning the Flames of Hate: Social Media and Hate Crime." *Journal of the European Economic Association* 19 (4): 2131–67.

Munn, Luke. 2019. "Alt-Right Pipeline: Individual Journeys to Extremism Online." *First Monday* 24 (6). https://doi.org/10.5210/fm.v24i6.10108.

Nozza, Debora, Federico Bianchi, and Dirk Hovy. 2020. "What the [Mask]? Making Sense of Language-Specific BERT Models." Working Paper.

Oktavianus, Jeffry, Brenna Davidson, and Lu Guan. 2021. "Framing and Counter-framing in Online Collective Actions: The Case of LGBT Protests in a Muslim Nation." *Information, Communication & Society* 26 (3): 479–95.

Perugorría, Ignacia, and Benjamín Tejerina. 2013. "Politics of the Encounter: Cognition, Emotions, and Networks in the Spanish 15M." *Current Sociology* 61 (4): 424–42.

Postmes, Tom, Nyla R. Branscombe, Russell Spears, and Heather Young. 1999. "Comparative Processes in Personal and Group Judgments: Resolving the Discrepancy." *Journal of Personality and Social Psychology* 76 (2): 320.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. "Language Models Are Unsupervised Multitask Learners." *OpenAI blog*.

Ragas, Matthew W., Hai L. Tran, and Jason A. Martin. 2014. "Media-Induced or Search-Driven? A Study of Online Agenda-Setting Effects during the BP Oil Disaster." *Journalism Studies* 15 (1): 48–63.

Rathje, Steve, Jay J. Van Bavel, and Sander Van Der Linden. 2021. "Out-Group Animosity Drives Engagement on Social Media." *Proceedings of the National Academy of Sciences* 118 (26): e2024292118.

Reid, Shannon E., Matthew Valasik, and Arunkumar Bagavathi. 2020. "Examining the Physical Manifestation of Alt-Right Gangs: From Online Trolling to Street Fighting." In *Gangs in the Era of Internet and Social Media*, eds. Chris Melde and Frank Weerman, 105–34. Cham, Switzerland: Springer.

Schmidt, Anna, and Michael Wiegand. 2019. "A Survey on Hate Speech Detection Using Natural Language Processing." In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, 1–10. Valencia, Spain: Association for Computational Linguistics.

Schulze, Heidi, Julian Hohner, Simon Greipl, Maximilian Girgnhuber, Isabell Desta, and Diana Rieger. 2022. "Far-Right Conspiracy Groups on Fringe Platforms: A Longitudinal Analysis of Radicalization Dynamics on Telegram." *Convergence: The International Journal of Research into New Media Technologies* 28 (4): 1103–26.

Scrivens, Ryan, Garth Davies, and Richard Frank. 2020. "Measuring the Evolution of Radical Right-Wing Posting Behaviors Online." *Deviant Behavior* 41 (2): 216–32.

Shirky, Clay. 2008. *Here Comes Everybody: The Power of Organizing without Organizations*. London: Penguin.

Simon, Bernd, and Bert Klandermans. 2001. "Politicized Collective Identity: A Social Psychological Analysis." *American Psychologist* 56 (4): 319.

Skitka, Linda J., Christopher W. Bauman, and Edward G. Sargis. 2005. "Moral Conviction: Another Contributor to Attitude Strength or Something More?" *Journal of Personality and Social Psychology* 88 (6): 895–917.

Skitka, Linda J., and G. Scott Morgan. 2014. "The Social and Political Implications of Moral Conviction." *Political Psychology* 35(S1): 95–110.

Skoric, Marko M., Qinfeng Zhu, Debbie Goh, and Natalie Pang. 2016. "Social Media and Citizen Engagement: A Meta-Analytic Review." *New Media and Society* 18 (9): 1817–39.

Snow, David, and Robert D. Benford. 1988. "Ideology, Frame Resonance, and Participant Mobilization." *International Social Movement Research* 1 (1): 197–217.

Su, Yan, Jun Hu, and Danielle Ka LaiLee. 2020. "Delineating the Transnational Network Agenda-Setting Model of Mainstream Newspapers and Twitter: A Machine-Learning Approach." *Journalism Studies* 21 (15): 2113–34.

Suler, John. 2004. "The Online Disinhibition Effect." *Cyberpsychology & Behavior* 7 (3): 321–6.

Tajfel, Henri, and John C. Turner. 2004. "The Social Identity Theory of Intergroup Behavior." In *Political Psychology: Key Readings*, eds. John T. Jost and Jim Sidanius, 276–93. New York: Psychology Press.

Takeshi, Amemiya. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Urman, Aleksandra, and Stefan Katz. 2022. "What they Do in the Shadows: Examining the Far-Right Networks on Telegram." *Information, Communication & Society* 25 (7): 904–23.

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press

Valenzuela, Sebastián, Teresa Correa, and Homero Gil de Zúñiga. 2018. "Ties, Likes, and Tweets: Using Strong and Weak Ties to Explain Differences in Protest Participation across Facebook and Twitter Use." *Political Communication* 35(1): 117–34.

Van Zomeren, Martijn, Tom Postmes, and Russell Spears. 2008. "Toward an Integrative Social Identity Model of Collective Action: A Quantitative Research Synthesis of Three Socio-Psychological Perspectives." *Psychological Bulletin* 134 (4): 504.

Velasquez, Alcides, and Gretchen Montgomery. 2020. "Social Media Expression as a Collective Strategy: How Perceptions of Discrimination and Group Status Shape U.S. Latinos' Online Discussions of Immigration." *Social Media + Society* 6 (1): 2056305120914009.

Vidgen, Bertie, Alex Harris, Dong Nguyen, Rebekah Tromble, Scott Hale, and Helen Margetts. 2019. "Challenges and Frontiers in Abusive Content Detection." In *Proceedings of the Third Workshop on Abusive Language Online*, 83–90. Florence, Italy: Association for Computational Linguistics.

Vidgen, Bertie, Dong Nguyen, Helen Margetts, Patricia Rossini, and Rebekah Tromble. 2021. Introducing CAD: the Contextual Abuse Dataset. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2289–303. Online, Association for Computational Linguistics.

Walker, Hunter. 2022. "Exclusive: January 6 Committee' Locked In' on Proud Boys." *Rolling Stone*, March 26. https://www.rollingstone.com/politics/politics-news/jan-6-committee-oath-keepers-proud-boys-first-amendment-praetorian-1327050/.

Walther, Samantha, and Andrew McCoy. 2021. "U.S. Extremism on Telegram." *Perspectives on Terrorism* 15 (2): 100–24.

Wang, Rong, Wenlin Liu, and Shuyang Gao. 2016. "Hashtags and Information Virality in Networked Social Movement: Examining Hashtag Co-Occurrence Patterns." *Online Information Review* 40 (7): 850–66.

Waseem, Zeerak, and Dirk Hovy. 2016. "Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter." In *Proceedings of the NAACL Student Research Workshop*, 88–93. San Diego, CA: Association for Computational Linguistics.

Williams, Matthew L., Pete Burnap, Amir Javed, Han Liu, and Sefa Ozalp. 2020. "Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime." *The British Journal of Criminology* 60 (1): 93–117.

Williamson, Vanessa. and Isabella Gelfand. 2019. "Trump and Racism: What do the Data Say?" *Brookings*, August 14. https://www.brookings.edu/blog/fixgov/2019/08/14/trump-and-racism-what-do-the-data-say/.

Wilson, Jason. 2018. "FBI Now Classifies Far-Right Proud Boys as 'Extremist Group,' Documents Say." *The Guardian*, November 18. https://www.theguardian.com/world/2018/nov/19/proud-boys-fbi-classification-extremist-group-white-nationalism-report

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach*. Boston, MA: Cengage Learning.

Zhong, Wei, Catie Bailard, David Broniatowski, and Rebekah Tromble. 2024. "Proud Boys on Telegram." *Journal of Quantitative Description: Digital Media* 4: 1–47.

https://doi.org/10.1017/S0003055423001478 Published online by Cambridge University Press