Countering Violent Extremism and Radical Rhetoric

Tamar Mitts

Abstract How do extremist sympathizers respond to counter-radicalization efforts? Over the past decade, programs to counter violent extremism have mushroomed around the world, but little is known of their effectiveness. This study uses social media data to examine how counter-radicalization efforts shape engagement with extremist groups in the online world. Matching geolocated Twitter data on Islamic State sympathizers with granular information on counter-extremism activities in the United States, I find that, rather than deradicalizing, these efforts led Islamic State sympathizers to act strategically to avoid detection. After counter-extremism activities, the group's supporters on Twitter who were in the vicinity of these events began self-censoring expressions of support for the Islamic State, altered profile images and screen names, and encouraged followers to migrate to Telegram, an encrypted network not viewable by the public. These findings reveal previously unknown patterns in the effects of counter-extremism programs in the digital era.

On 3 June 2017, a British man drove a truck into a crowd of pedestrians on the London Bridge before going on a stabbing rampage in a nearby market. According to friends and neighbors, the man had been inspired to carry out the violence by watching extremist videos on YouTube.[1] On 27 October 2018, another man stormed into a synagogue in Pittsburgh, Pennsylvania, shooting and killing eleven people during weekend prayers. Shortly before the attack, he had written on social media of his fight against the Jews for facilitating an influx of migrants into the United States. Although he acted alone, investigators believe that he was radicalized by engaging with extremist content online.[2]

These stories illustrate a growing phenomenon that has gripped the attention of many governments in recent years. The digital revolution and the rapid expansion of information and communication technologies have made it easier for terrorist groups to recruit supporters and inspire violence around the world. Groups espousing extremist ideologies are now actively using social media platforms to disseminate and promote their ideas.[3] The Islamic State (IS), for example, recruited thousands of individuals around the world by disseminating militant propaganda through elaborate

1. Steven Erlanger, "Another Terrorist Attack Strikes the Heart of London," *New York Times*, 3 June 2017.

2. Kevin Roose, "On Gab, an Extremist-Friendly Site, Pittsburgh Shooting Suspect Aired His Hatred in Full," *New York Times*, 28 October 2018.

3. Berger 2015; Hamm, Spaaij, and Cottee 2017; Mitts 2019.

International Organization 76, Winter 2022, pp. 251–72

doi:10.1017/S0020818321000242

campaigns on Facebook, Twitter, YouTube, and Telegram. Similar trends were seen with far-right extremist websites which inspired terrorism by promoting hate against minorities.

To combat radicalization, governments began enacting policies to regulate social media content and prevent hate speech, misinformation, and violent propaganda from influencing their citizens. Some focused on challenging extremist ideologies by flooding social networking sites with counter-speech campaigns. The US State Department, for example, led a campaign to encourage individuals sympathetic to IS to "think again" and "turn away" from the group.[4] Another approach, led by technology companies, focused on taking down content promoting violence—an initiative that resulted in the suspension of millions of accounts from various social media platforms in just a few years.[5]

But efforts to counter violent extremism extend beyond online content. Many governments looked for ways to prevent extremists from inspiring violence in the "offline" world too. At the forefront of these initiatives were community engagement programs, which sought to encourage citizens to serve as "early warning systems" by sharing information with the government on individuals who might pose security risks. The idea was that citizens with personal connections to radicalizing individuals, like friends or family members, are best positioned to detect changes in behavior that might convey early signs of radicalization.[6] By raising citizen awareness of violent extremism, governments hoped to access crucial information on potentially radicalizing individuals to monitor their behavior and, if needed, stop them from engaging in violence.[7]

Despite these increased efforts to combat extremism, we know very little about what happens in the online world when counter-extremism efforts take place on the ground. There is a large literature on how counter-radicalization programs affect targeted communities in areas other than extremism.[8] There is also a growing body of research on the effects of counter-speech campaigns on online engagement with violent groups.[9] However, there is little empirical work on how "offline" efforts to counter extremism shape the online behavior of those sympathetic to violent extremist groups.

This study provides the first large-scale, systematic analysis of the effects of on-the-ground counter-extremism initiatives on engagement with terrorist groups in

4. Fernandez 2015.

5. Davey Alba, Catie Edmondson, and Mike Isaac, "Facebook Expands Definition of Terrorist Organizations to Limit Extremism," *New York Times* 19 September 2019; Mike Isaac, "Twitter Steps Up Efforts to Thwart Terrorists' Tweets," *New York Times*, 5 February 2016.

6. Briggs 2010; Dalgaard-Nielsen and Schack 2016; Dunn et al. 2016; Romaniuk 2015; Vermeulen 2014.

7. One of the major controversies with this approach is that these interventions often occur before individuals commit a crime, which can encroach on their constitutionally protected rights. For a discussion, see Jackson et al. 2019.

8. Gillum 2018; Thomas 2010.

9. Davey, Birdwell, and Skellett 2018; Helmus and Klein 2018.

the online world. I focus on a large program to prevent radicalization in the United States, which was initiated by the Obama administration to counter *jihadi*-inspired terrorism, and examine how it shaped the behavior of IS sympathizers on Twitter.

I build on a large body of research on terrorism, insurgency, and online mobilization to test two mechanisms by which counter-extremism efforts might shape online engagement with violent groups: deradicalization and strategic behavior. *Deradicalization* refers to individuals with extreme worldviews adopting more moderate positions after exposure to counter-extremism interventions. In the context of support for IS, this might be observed as a decline in rhetoric supporting the group, its ideology, and its actions around the world, as well as disengagement from online networks affiliated with it. *Strategic behavior* refers to extremist supporters becoming aware of government surveillance and altering their behavior to avoid detection. This might be observed as attempts to alter online identities or migration to private social media platforms that are less observable by the public.

Drawing on rich Twitter data on IS sympathizers in the United States and granular information on community engagement events taking place between 2014 and 2016, I examine how counter-extremism interventions affected online engagement with IS on social media. I identified the timing and location of dozens of community engagement activities taking place across the United States and matched them with geolocated data generated by IS sympathizers on Twitter: several million tweets with rich user-profile metadata. I study changes in online pro-IS rhetoric, as well as other behaviors that are indicative of awareness of surveillance, such as profile-picture or screen-name changes and the promotion of encrypted online platforms. Using difference-in-differences models, I examine how IS sympathizers behaved online after these events took place in their vicinity.

I find that community engagement activities, which encouraged citizens to "say something" if they "see something," led IS sympathizers on Twitter to engage in strategic behavior. In the weeks following counter-extremism events, the group's supporters in event areas significantly reduced the number of tweets endorsing IS, including posts expressing sympathy with the group, describing life in IS-controlled territories, and mentioning foreign fighters. While this decrease could be interpreted as evidence of deradicalization, I find that proximity to counter-extremism activities led IS sympathizers to take additional actions that indicate their continued support for the group and their awareness of surveillance. After counter-extremism events, these users altered their public identity on Twitter by changing their profile pictures and screen names, and increased the number of propaganda-disseminating accounts they followed on the platform. When the group began migrating to Telegram in the end of 2015, these users advertised the new platform to their followers, suggesting that they may have switched to Telegram to avoid detection.[10]

10. Telegram is an encrypted platform that became popular among IS supporters on Twitter, especially after their activities on public, mainstream platforms were disrupted; see Bloom, Tiflati, and Horgan 2019.

These results challenge the assumptions underlying many counter-extremism interventions, which presume that citizen vigilance can help sway at-risk individuals from the path of violence. By showing how government-sponsored community engagement activities propelled IS sympathizers to behave strategically on social media, the study points to an underappreciated consequence of these emerging efforts. As violent groups are increasingly using online platforms for recruitment, it is important to understand how on-the-ground activities to prevent radicalization shape the online behavior of extremist supporters. After all, the next perpetrator of a terrorist act might be inspired by interacting with these groups online.

Data Sources

Counter-Extremism Events in the United States

In August 2011, the Obama administration initiated a counter-radicalization strategy, Empowering Local Partners to Prevent Violent Extremism in the United States, which focused on strengthening the government's engagement with local communities whose members may be targeted by violent groups.[11] While the plan's official goal was to target violent extremism across all ideologies, in practice, the vast majority of its activities focused on *jihadi*-inspired extremism—an approach that generated much opposition among civil rights activists.[12]

A large part of these efforts consisted of community engagement events, which sought to increase citizen awareness of the threat of extremism and encourage collaboration with the government to prevent violence. These activities included, for example, community roundtables that brought together government officials and members of local communities to strengthen relationships and share information, and community awareness briefings, in which government officials presented details on the process of radicalization and online recruitment by terrorist groups.[13] Figure 1 shows advertisements of community engagement events in New York and Colorado, and Figure 2 presents photos from activities in Georgia and Arizona.

Although these events attracted mostly community leaders and some interested citizens, information on the government's call to "see something, say something" was shared throughout the community. Individuals in areas that were targeted by counter-extremism activities were encouraged to keep an eye out for suspicious

11. See <https://www.dhs.gov/sites/default/files/publications/empowering_local_partners.pdf>.

12. Critics argued that the targeting of Muslim communities with counter-extremism programming was discriminatory and dangerous, both stereotyping Muslims as "security threats" and engendering a climate of fear that discouraged the free expression of political opinions. In addition, many argued that the focus on Muslim communities is unjust, as most casualties since 9/11 have been caused by far-right terrorism. See American Civil Liberties Union 2016; Council on American-Islamic Relations 2016; Gillum 2018; Kundnani 2009; Patel and German 2015.

13. The Department of Homeland Security has been leading these efforts in recent years. Many community engagement meetings have not exclusively focused on extremism but covered a wide range of issues related to the department's activities.

FIGURE 1. *Advertisements for community engagement events aimed at increasing local awareness of violent extremism (data from CrowdTangle, a public insights tool owned and operated by Facebook)*

FIGURE 2. *Photos from community engagement events in Atlanta, GA, and Phoenix, AZ (photos courtesy of Islamic Speakers Bureau, Atlanta, and Islamic Community Center of Phoenix)*

behavior, even if they did not attend these meetings. Take the case of Sal Shafi. In response to the government's call for vigilance, he decided to report his son's behavior to the FBI after observing him consume extremist content online, changing his appearance, and trying to travel to Turkey.[14] Many family members in a similar

14. Matt Apuzzo, "Only Hard Choices for Parents Whose Children Flirt With Terror," The New York Times, 9 April 2016; Nate Gartrell, "In Rebuke of Feds, Judge Frees East Bay Man Once Accused of Terrorism," The Mercury News, 30 March 2019.

TABLE 1. *Twitter posts by IS sympathizers in the United States, sharing information on community engagement events*

RT @NYPDMuslim: #HappeningNow #PBBX holding its Pre #Ramadan meeting w/
#Bronx #Muslim Community Leaders thank u 4 the great support
#VIDEO Community Engagement Night with the Department of Homeland Security [URL] #rumiforum
Awaiting Homeland Sec. Jeh Johnson at All Dulles Area Mosque in Va. [URL]
The #DHS says the social media platform is a "constant provider [of intelligence] and is fairly reliable"

situation informed government authorities of their relatives' behavior out of a desire to prevent violence.[15]

Local communities were also encouraged to monitor their members' online behavior, as many extremist sympathizers maintained public profiles on social media.[16] Amani Ibrahim, for example, tried to stop her teenage son from going on the Internet after becoming aware of his advocacy for IS on Twitter. When her efforts were not successful, she followed the advice of a community leader and reported her son's behavior to the government, which resulted in him being arrested and sentenced to several years in prison.[17] Indeed, many anti-terrorism indictments against American citizens drew heavily on information shared on social media.[18] Thus, even though individuals who engaged with IS online did not attend community engagement events, they were likely aware of them; some even shared information on such events and talked about the government's monitoring of social media with their Twitter followers (Table 1).

I collected information on community engagement events taking place between 2014 and 2016 from newsletter reports published by the US government.[19] I gathered data on the dates of these events, the cities in which they took place, and the type of engagement activity carried out in each event. Figure 3 shows the number of community engagement events by month. The online supplement provides more information on these activities and detailed summary statistics tables.

Islamic State Sympathizers on Twitter

To assess whether these activities influenced the behavior of extremist supporters online, I use Twitter data on IS sympathizers in the United States who interacted

15. Kristina Cooke and Joseph Ax, "US Officials Say American Muslims Do Report Extremist Threats," Reuters, 16 June 2016.
16. Department of Homeland Security 2015.
17. Warren Richey, "One Virginia Teen's Journey from ISIS Rock Star to Incarceration," *Christian Science Monitor*, 29 September 2015.
18. Greenberg 2016.
19. The newsletters were published by the Office of Civil Rights and Civil Liberties in the Department of Homeland Security. Figure A1 in the online supplement shows an example of a newsletter report from August 2015.

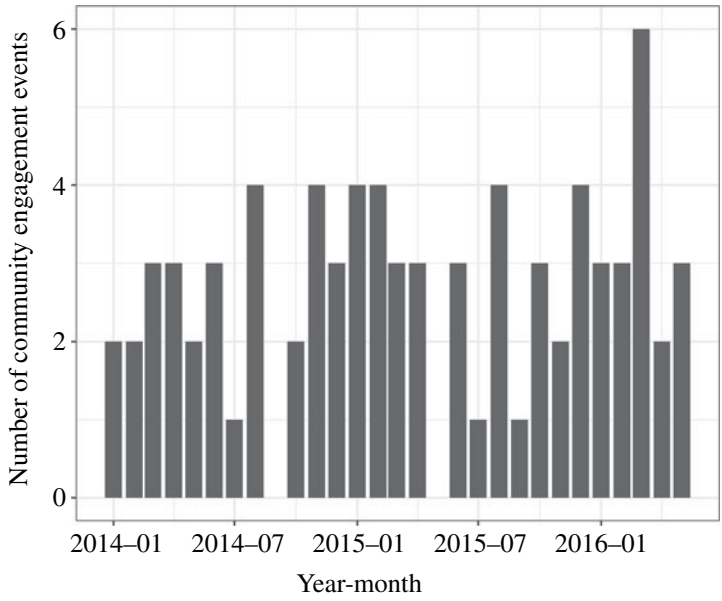

FIGURE 3. *Community engagement to counter extremism: monthly number of community roundtables held by the Department of Homeland Security in the United States from 2014 to 2016*

with the group on the platform between 2014 and 2016. I provide an overview of the data collection process, which includes (1) identifying accounts of IS supporters on Twitter and downloading information on their posting history; (2) coding the extent to which their posts reflect support for IS; (3) measuring changes in their public Twitter profiles; (4) predicting users' locations; and (5) geographically matching users with community engagement events. More details on each of these steps are in the online supplement.

1. Identifying IS accounts on Twitter. Using Twitter's public APIs,[20] I collected detailed information on accounts of IS sympathizers on Twitter who followed at least one of about 15,000 accounts that actively disseminated IS propaganda on the platform.[21] These users engaged with the group's Twitter networks in various capacities. Some expressed strong sympathy with the group and its ideology; some actively "retweeted" its propaganda, while others were more passive in their online engagement with the group. I downloaded every available piece of information on these

20. Application programming interfaces (APIs) allow systematic retrospective and prospective data collection from websites.

21. I identified propaganda-disseminating accounts by live-tracking "black lists" published by anti-IS hacking groups who monitored the group's activity on Twitter. See the online supplement for details.

accounts before they were suspended from the platform, including data on user profiles, screen names, locations, historical tweet timelines, and lists of friends and followers. In total, I collected information on 30,358 users in the United States, who posted a total of 15,140,867 tweets between 2014 and 2016.

2. Measuring online expression of extremist ideology. Using the historical tweet timelines of these accounts, I measured the extent to which each post represented pro-IS content. Since the volume of tweets was large, I used supervised machine learning to classify tweets into different content categories: expressions of sympathy with the group, discourse on life in IS-controlled territories, communications about the group's actions in the Syrian civil war, and mentions of foreign fighters. I used a crowdsourcing platform to manually label a training set of about 30,000 randomly selected posts, and trained models to predict the content of unlabeled tweets in each category.[22] I created a tweet-level index variable capturing pro-IS sentiment across all four categories. The index was constructed by summing the predicted content scores of each tweet along the four categories, and normalizing the sum to range between 0 and 1. To make the tweet-level data easy to interpret at the user level, I generated an aggregated variable that measures this content in each user's Twitter posts in the week before and one to four weeks after community engagement events.[23] Table A9 in the online supplement shows summary statistics for this variable.

3. Measuring changes in profile-level metadata. In addition to content, I measured other online actions that individuals might take to evade surveillance. Drawing on user-level metadata provided by Twitter's public APIs, I collected information on IS sympathizers' profile pictures and screen names, and the number of propaganda-disseminating accounts they followed on the platform. This includes weekly observations of IS sympathizers' profile information, sampled every seven days, from January to June of 2016. I created a user-week-level data set measuring changes in these variables in the week before and one to four weeks after community engagement events, and used it to examine whether those users who were near counter-extremism activities were more likely to alter their online identities in their aftermath.[24] Section 4 of the online supplement provides details on the user metadata. Summary statistics show that the vast majority of IS sympathizers did not change their profile pictures or screen names during this period (Table A10 in the online supplement).

22. Section 2 of the online supplement provides details on the definitions of the content categories, the content analysis method, and model performance.

23. In the "pre" period, it is the average of the tweet-level index for each user in the week before community engagement events. In the "post" periods, it is the average of the index in each time window: first week, two weeks, three weeks, or four weeks after community engagement events. Section 2 of the online supplement provides more details on the creation of the index variable.

24. For example, when a user changed his or her profile picture in a given week, the *change profile picture* variable was coded as 1 for that user in that week; otherwise it was coded as 0. In the difference-in-differences analysis, I pool the data for each time window: first week, two weeks, three weeks, and four weeks after community engagement events.

4. Predicting IS supporters' locations. Since few social media users enable geotagging of their posts or provide location information in their accounts,[25] I estimated the locations of IS sympathizers on Twitter with an algorithm that predicts user locations from geographic information available in their networks.[26] In several contexts, this algorithm has been able to predict geolocation relatively accurately, outperforming other methods that rely on network data.[27] I use predicted locations in this study to avoid relying on the small subset of users with reported locations. Section 3 of the online supplement provides details on the method, along with information on prediction accuracy and model stability.

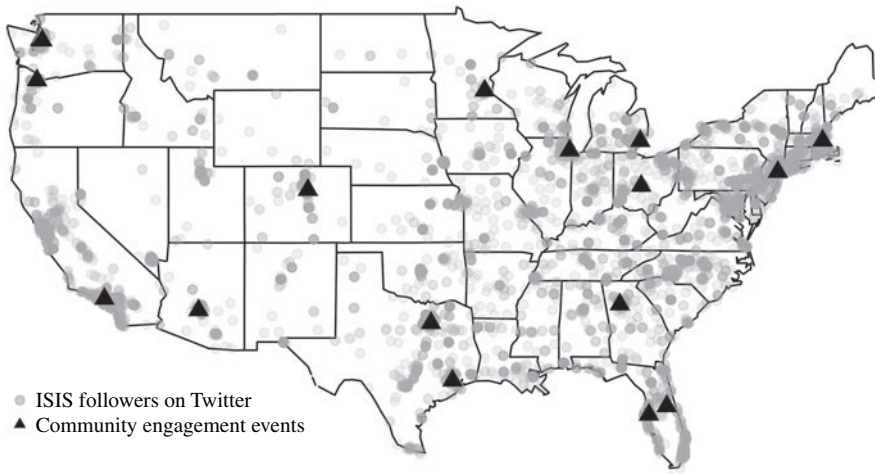

FIGURE 4. *Estimated US locations of Islamic State sympathizers on Twitter, mapped along with community engagement events of the Office of Civil Rights and Civil Liberties*

5. Matching IS sympathizers to community engagement events. To determine which users were near community engagement events, I matched users' predicted coordinates with geographic information on localities across the United States. I used the US Census TIGER database to access spatial data on US localities (cities, towns, etc.), and coded each location for whether it was targeted by counter-extremism programming.[28] Figure 4 displays the predicted locations of IS sympathizers on

25. Recent studies estimate that only 2 to 3 percent of Twitter posts include location information. Leetaru et al. 2013.

26. Spatial label propagation algorithms rely on the finding in social network research that location information in a user's online network is a powerful predictor of a user's offline geographic location.

27. Jurgens et al. 2015.

28. The matching was done by overlaying coordinate data with the TIGER shape files (see <https://tinyurl.com/y2sqnvld> for more information). For robustness, I also use a measure of the distance in

Twitter, along with the locations of community engagement events taking place between 2014 and 2016. In my analysis, I compare the online behavior of IS sympathizers who were in event areas to those who were not, before and after each event.

Empirical Strategy

To analyze the relationship between community engagement activities and the online behavior of IS sympathizers on Twitter, I measured IS sympathizers' online behaviors before and after each event, which included information on the tweets they posted, whether they changed their profile pictures or screen names, and the number of propaganda-disseminating handles they followed. For each individual in each event, I created binary indicators to distinguish between (1) observations appearing before and after each event and (2) individuals inside or outside the area of the event.[29] Specifically, for each community engagement activity, I created the variable POST, which is coded 1 when an online action took place after the event and 0 otherwise, and the variable IN EVENT AREA, which is coded 1 when the action was taken by an individual in the area of the event and 0 otherwise.

Since I study the behavior of IS sympathizers around dozens of community engagement events, I conduct a pooled difference-in-differences analysis where I examine all community engagement events simultaneously. For each of these outcomes,

I estimate the following ordinary least squares regression:[30]

$$y_{i,j,k} = \beta_1 \text{POST}_{i,k} + \beta_2 \text{ IN EVENT AREA}_{j,k} + \beta_3(\text{POST}_{i,k} \times \text{IN EVENT AREA}_{j,k}) + \alpha_k + \varepsilon_j \quad (1)$$

where $y_{i,j,k}$ is an online action i (expressing pro-IS rhetoric, changing profile pictures or screen names, or following/unfollowing of propaganda-disseminating accounts) by user j surrounding event k; $\text{POST}_{i,k}$ is 0 when the action took place before event k, and 1 afterwards; $\text{IN EVENT AREA}_{j,k}$ is 1 when the action was taken by an individual in the area of event k, and 0 otherwise; and α_k is an event fixed effect. In all specifications, β_3 is the difference-in-differences coefficient of interest, reflecting how the online behavior of IS sympathizers after community engagement events is different between individuals who were in an event area and those who were not.[31] Standard errors are clustered at the user level.

kilometers of each user from the center of the city/town in which community engagement events took place. The results remain the same (Figure A7 in the online supplement).

29. I define the "event area" as the geographic boundaries of the locality in which a community engagement event was held. The results do not change when using distance in kilometers as a measure of proximity (Figure A7 in the online supplement).

30. The research note presents OLS results for all outcomes, but the online supplement (Section 6.3) presents additional analyses using logit and Poisson regressions for binary and count outcomes, respectively.

31. To account for prediction error in the pro-IS rhetoric variable and the location of users, both of which were estimated from machine learning models, I added weights to all regressions that give higher weight to observations with smaller errors. The results are very similar if I do not include these weights.

The key identifying assumption is that in the absence of a community engagement event, individuals who are in the event area and individuals who are not follow parallel trends in their online support for IS. While it is certainly likely that the US government targeted specific areas that it deemed more likely to have individuals "at risk" of radicalization,[32] changes over time in support for IS should not be significantly different between the groups before the occurrence of community engagement events at the high frequency (i.e., a few days).

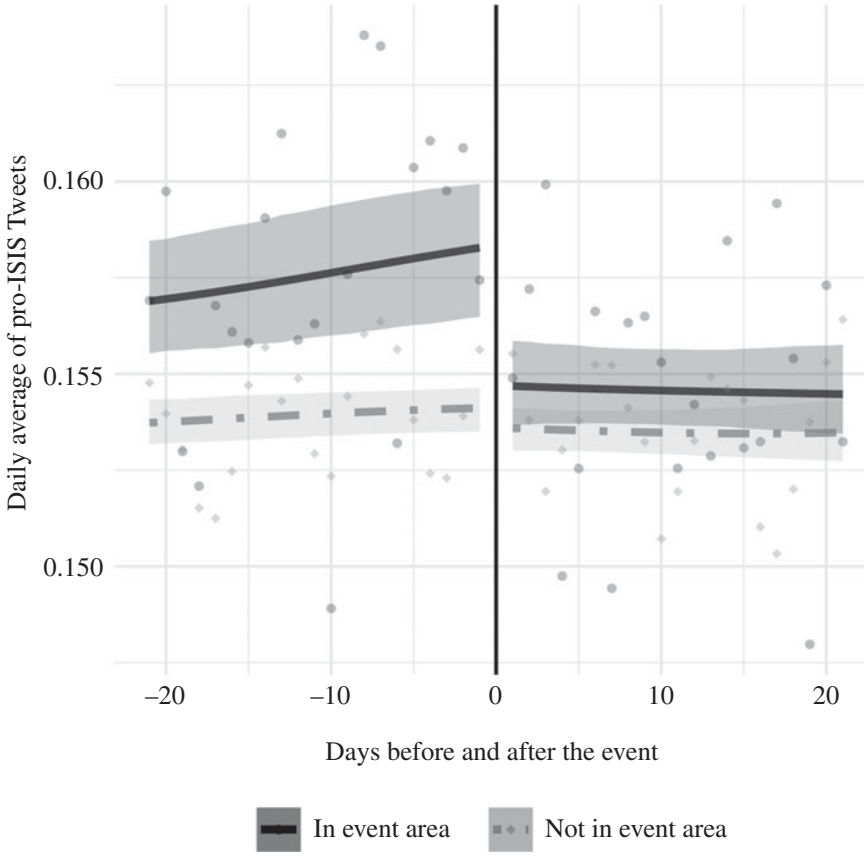

FIGURE 5. *Pro-IS rhetoric: parallel trends*

To empirically test this assumption, I visually examine whether the two groups display parallel trends before community engagement events. Figure 5 plots pre- and post-time trends in pro-IS rhetoric for the individuals who were in event areas

32. Bjelopera 2014.

(black) and those who were not (gray). The *x*-axis is the number of days between a community engagement event and when IS followers posted on Twitter. To observe time trends for all events simultaneously, I normalized the difference in days between the events and the timing of Twitter posts.

The time trends in pro-IS rhetoric are parallel in the pretreatment period. Only after community engagement events do we observe a shift in those trends, where pro-IS content by individuals in event areas decreases, but the rhetoric of those outside of event areas does not change. We also observe that the average pro-IS rhetoric before community engagement events is higher for individuals in event areas. Table A11 in the online supplement presents a statistical test of the parallel-trends assumption; there is no difference in the time trends between the groups, except when expanding the pretreatment data back to thirty days before the events.

Results

I first present results for content produced by IS sympathizers on Twitter, showing that community engagement events led to a decrease in pro-IS rhetoric. I then present evidence suggesting that this decrease is likely driven by strategic behavior, by showing that individuals near community engagement events also changed their profile pictures and screen names, and increased their following of propaganda-disseminating accounts. Finally, I present results suggesting that these users also sought to avoid detection by migrating to Telegram, an encrypted online platform that became popular among IS supporters during the years of this study.

Changing content: a decrease in pro-IS rhetoric. Panel A of Figure 6 reports the findings for IS sympathizers' pro-IS rhetoric on Twitter. The figure plots coefficients on the interaction term, POST IN EVENT AREA, estimated from regressions where the dependent variable is measured at different time intervals in the post-treatment period, ranging from one to thirty days.[33] I find that community engagement activities reduced expressions of support for IS among users in event areas. In almost all specifications, the difference-in-differences coefficient is estimated at 0.005. By way of reference, the pretreatment difference between users inside and outside the event areas is 0.006, so the magnitude of the estimated effect is economically meaningful. Since this study examines a period when IS was actively expanding its territorial control and publicizing its cause on the Internet, this decrease is notable.

Strategic online behavior. Since a decrease in online expressions of support for IS can be interpreted as evidence of deradicalization but also as strategic behavior, I next examine additional actions that are more likely to be associated with strategic

33. Full results are shown in Table A12 in the online supplement.

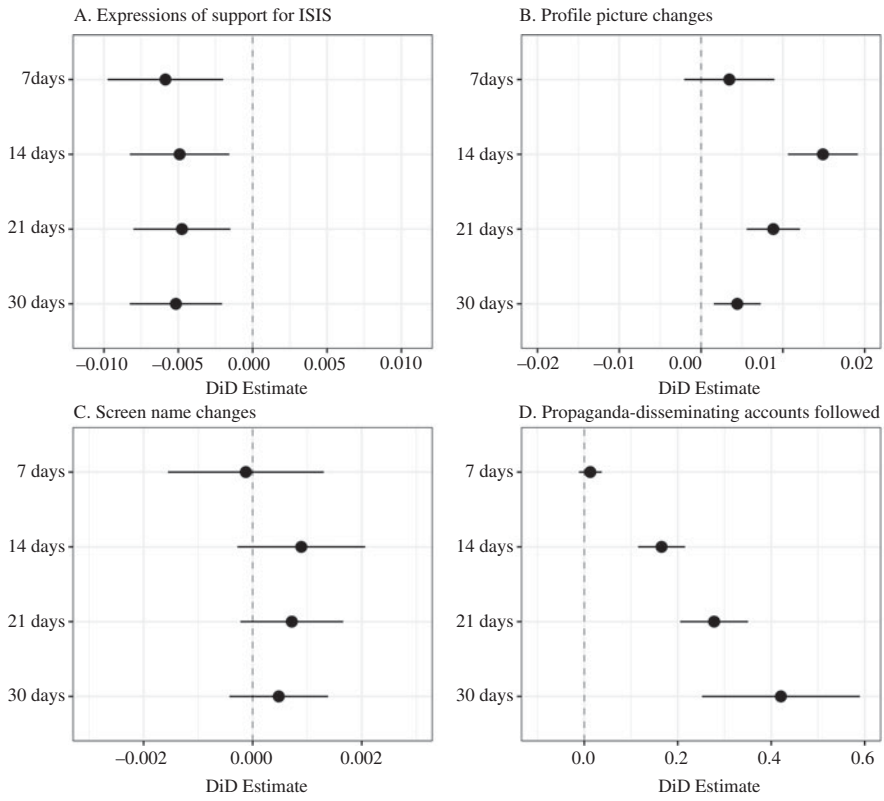

FIGURE 6. *The impact of community engagement on Islamic State sympathizers on Twitter*

behavior. If these behaviors take place alongside a decline in pro-IS rhetoric, then it would be reasonable to infer that the decrease in online expressions of support for IS is not driven by deradicalization but by awareness of surveillance in the wake of community engagement events.

In Panels B and C of Figure 6, I examine whether community engagement events propelled IS sympathizers to change their profile pictures and screen names on Twitter. The figures present difference-in-differences coefficients estimated from regressions in which these outcomes are measured across different time windows, from one to four weeks before and after the events.[34] I find that IS sympathizers were significantly more likely to change their profile pictures and screen names after community engagement activities took place in their areas, a trend that was

34. Tables A13 and A14 in the online supplement present the full results. The findings do not change when using logit regressions; see Tables A21 and A22.

strongest about two to three weeks after the events.[35] Since altering Twitter profiles was very infrequent (about 3 percent of the sample), the concentration of these changes in areas with counter-extremism activities is revealing.

Panel D shows that IS sympathizers in event areas also increased the number of propaganda-disseminating accounts they followed on Twitter.[36] When compared to the pretreatment difference between users inside and outside event areas, the change reflects about a 30-to-55-percent increase in the number of Twitter handles disseminating IS propaganda. This cuts against the deradicalization interpretation because even though IS sympathizers in event areas refrained from publicly endorsing the group on the platform, they continued to passively engage with the group's Twitter networks by following more accounts.

To check that these changes are in fact happening simultaneously, I next examine whether reducing pro-IS rhetoric, changing profile pictures, changing screen names, and following propaganda-disseminating accounts are done by the same users, as opposed to a mix of different users, in event areas. In the analysis, each of these actions was examined separately. It is possible that users who reduced their pro-IS rhetoric did not also change their profile picture, and vice versa, even though in the aggregate this happened more often in event areas. Figure 7 presents difference-in-differences coefficients from regressions where the dependent variables (reported in the rows) reflect different combinations of online actions.[37] I find that many users in event areas took two or more actions after counter-extremism events, and some even engaged in three or more.[38] This provides further support for the strategic-behavior interpretation.

Migration to Telegram. Finally, I examine whether IS sympathizers migrated to private communication channels after counter-extremism events. Telegram is a messaging application that became popular among IS supporters when their accounts were suspended from mainstream platforms like Twitter and Facebook.[39] In its early days, Telegram offered only one-to-one messaging, but in September 2015 it launched a "channels" service that enabled sharing content with an unlimited number of followers. Shortly after the launch, IS channels proliferated on the platform, making Telegram one of the organization's main online hubs in recent years.[40] Several major terrorist attacks were linked to engagement with IS recruiters

35. The reason the changes were more noisy within a seven-day window is that these outcomes were measured once a week. If a community engagement event occurred just after the sampling date, the effect would not appear until the next sampling date, which in this case would be the fourteen-day window. See Section 4 of the online supplement for details.

36. Table A15 in the online supplement presents the full results. Robustness tests using Poisson regressions are shown in Table A23.

37. For each combination, I created a dummy variable coded as 1 when a user took the actions reported in the row simultaneously, and 0 otherwise.

38. Table A16 in the online supplement shows full results.

39. Berger and Perez 2016.

40. Bloom, Tiflati, and Horgan 2019.

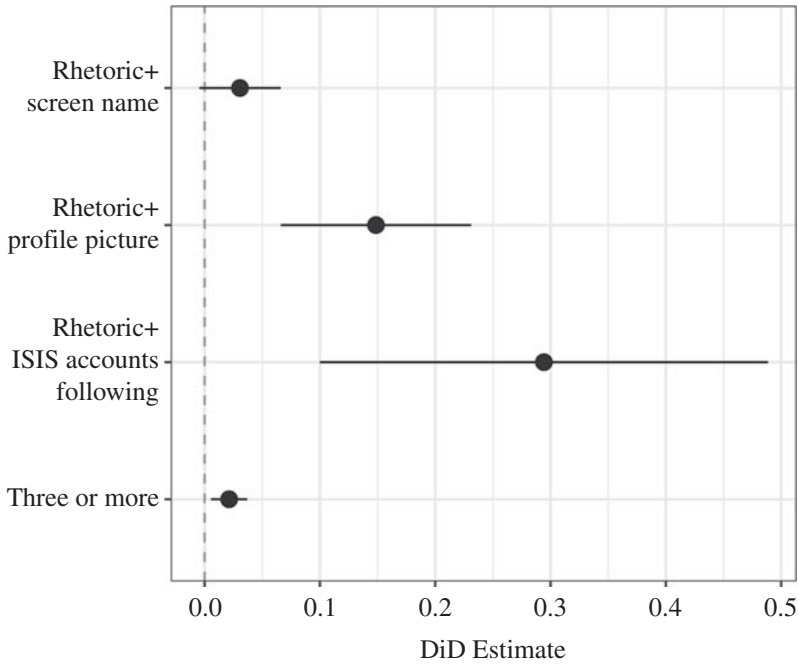

FIGURE 7. *Engaging in several online actions simultaneously*

on Telegram, including the attacks in Paris, Berlin, and Istanbul in 2015, 2016, and 2017.[41]

To assess whether IS sympathizers on Twitter became interested in Telegram after community engagement events, I ran several additional tests. First, I measured mentions of Telegram in these users' Twitter communications, by coding each tweet for whether it included the word *telegram*.[42] I find that about 3 percent of the users mentioned Telegram between 2014 and 2016, and that mentions of the platform became significantly more frequent after 22 September 2015, when Telegram launched its channels. Panel A of Figure 8 shows the frequency of IS sympathizers' mentions of Telegram on Twitter, and Table 2 shows some examples. It can be seen that much of the discourse on Telegram related to either the opening of new accounts on the platform, or the advertisement of IS Telegram channels.

Second, I examine whether mentions of Telegram among individuals in event areas were more numerous after community engagement events. Panel B of Figure 8 presents difference-in-differences coefficients, where the dependent variable is an indicator coded 1 if a user mentioned the word *telegram* in one or more tweets and 0

41. Rebecca Tab, "Terrorists' Love for Telegram, Explained," *Vox,* 30 June 2017; Yayla 2017.
42. I use English-language tweets for this purpose.

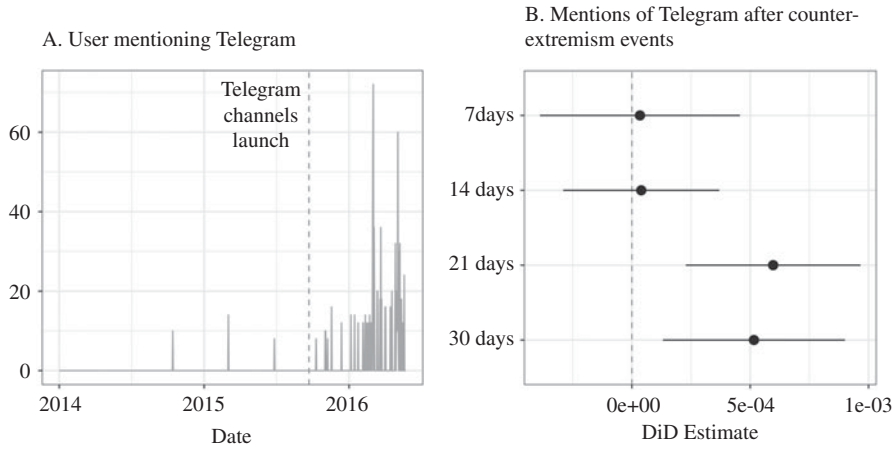

FIGURE 8. *Mentions of Telegram in Twitter posts. After Telegram launched its channels service, Islamic State sympathizers in event areas significantly increased their references to this alternative platform.*

otherwise. As before, I estimate models for different time windows, ranging from one to four weeks after the events. I find that IS sympathizers close to community engagement activities significantly increased their mentions of Telegram after these events. While the changes were less evident in the first two weeks, they became noticeable after three weeks.

TABLE 2. *Examples of English-language tweets mentioning Telegram, posted by Islamic State sympathizers in the United States*

Alhamdulillah you can now access many of Sheikh Omar Bakri's audios,videos & writings on the Telegram via @…
I accidentally made my Telegram private! Thank you guys and girls for the help with the app! I can't do it without you!
I will be getting a Telegram!
don't know what telegram is? I will search it and start using it!
RT @…: New Telegram channel for all Khilafah Materials (Videos,Pics,Audios,News,Books) [URL] #BreakingNews
@… Whats Your Telegram Username ??
@… is there a telegram handle
ATTENTION Follow this telegram channel which will Insha Allah be of a benefit to us all [URL]

Third, I test whether Twitter usage in event areas went down after Telegram became a viable alternative to Twitter. I counted the number of tweets each user posted in the weeks after community engagement events, and examined whether it differed between users inside and outside event areas after the launch of Telegram's channels on 22 September 2015. After the introduction of IS's Telegram channels, IS sympathizers in event areas posted fewer tweets after community engagement events than those

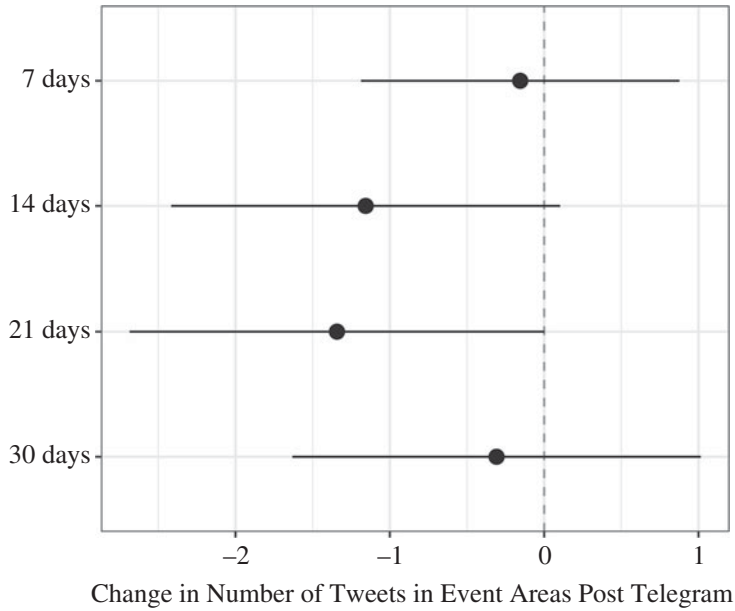

FIGURE 9. *After the introduction of IS's Telegram's channels, sympathizers close to community engagement activities posted fewer tweets in the aftermath of those events.*

outside the event areas (Figure 9). Coupled with the previous findings, this indicates that they might have migrated to Telegram to avoid detection.

Taken together, these results show that on-the-ground activities to counter violent extremism led to changes in the online behavior of IS sympathizers on Twitter. After community engagement events, these users reduced their online expressions of support for IS, took steps to change the appearance of their Twitter profiles, and promoted migration to Telegram, an alternative encrypted platform popular among the group's supporters.

Conclusion

What do these results imply for the risk of online-inspired terrorism? If encouraging vigilance in local communities motivates extremists to behave strategically, current counter-radicalization efforts may be ineffective, at least for those who actively interact with extremist groups online. Numerous cases have illustrated how engagement with terrorist content on social media can facilitate radicalization and violence.[43]

43. Carter, Maher, and Neumann 2014; Greenberg 2016; Vidino and Hughes 2015.

This study's findings suggest that counter-extremism programs might be pushing these individuals into closed online communities, which are less visible to the public and where the risk of radicalization may be higher. Indeed, the trade-off between counter-extremism activities and online backlash mirrors more general debates on the regulation of social media platforms. On the one hand, banning extremist content reduces the availability of violent ideologies on the Internet, potentially disrupting extremists' recruitment efforts. On the other hand, outright bans can motivate extremist groups to establish secret online networks that may be harder to monitor.

Many social media companies have chosen the content-takedown approach, even with the risk of pushing extremists into more radical corners of the Internet.[44] Facebook, for example, hired more staff to work on "dangerous organizations"—a specialized team of experts focused on designing policies and building algorithms to block violent content before it is uploaded. Google created tools to redirect searches for violent propaganda toward alternative content, while taking down millions of YouTube videos that promoted extremist violence. Twitter similarly stepped up its efforts to suspend accounts that disseminate violent propaganda, and joined broader initiatives in the tech sector to coordinate the blocking of extremist content across platforms. While evidence is still accumulating, several examples illustrate how content takedowns motivate migration to extremist communities on alternative online platforms.[45]

This study is the first to document a similar dynamic with "offline" efforts to prevent radicalization. Combining information on dozens of local activities to counter extremism with rich online Twitter data, I show that on-the-ground efforts motivated IS sympathizers to act strategically online, changing their public behavior on Twitter and promoting migration to Telegram. These findings highlight the growing importance of alternative online platforms for militant mobilization in the digital era. As extremists become aware of the "downsides" of public engagement on mainstream social media sites, they have a growing influence in smaller online communities—networks that are rarely systematically studied. Future research in this area should therefore consider not only mainstream social media platforms but also fringe, alternative online spaces.

I expect this study's findings to apply to other countries that have implemented similar counter-extremism strategies in recent years. The form of community engagement examined in this paper is very similar to activities taking place in the United Kingdom, Australia, Canada, and the Netherlands. I also expect to find similar dynamics among other groups using social media platforms, in particular white supremacy groups, which have dramatically increased their online mobilization in recent years.[46] If awareness of government surveillance pushes extremist supporters

44. Fishman 2019.
45. Bloom, Tiflati, and Horgan 2019; Nouri, Lorenzo-Dus, and Watkin 2019.
46. Marwick and Lewis 2017; Tech Transparency Project 2020.

to more private online platforms, the dynamics documented in this study are likely to be present in other communities as well.

But does all this imply that counter-extremism efforts are always ineffective? No. In this research note, I focused on a particular subset of the population, individuals who followed IS accounts on Twitter, and who often produced content that led to their ban from the platform. By examining this particular sample, I was able to shed light on how extreme individuals react to counter-radicalization initiatives. However, the results might not generalize to less radical individuals, for example, those who do not actively engage with IS social media but still find its ideology appealing. Indeed, several recent cases suggest that counter-radicalization programs can be effective at swaying more moderate individuals from the path of violence.[47] What this study shows is that it is important to pay attention to what extremists are doing in the online sphere when considering the impact of counter-radicalization efforts.

In addition, in this study I examine responses to counter-extremism activities within a few weeks after they occurred. While this allows me to study immediate reactions, it does not facilitate measuring the long-term consequences of these efforts. Deradicalization might take place over a longer period, even for "extreme" individuals who begin the process by self-censoring and migrating to encrypted platforms. The longer-term consequences of counter-radicalization programs are an important area for future research.

More broadly, this research contributes to a new wave of scholarship on the link between the digital revolution and the dynamics of conflict and violence.[48] In the last several years, numerous studies have illustrated how mobile phone technologies shape patterns of violence in civil war;[49] how social media facilitates mass protest and dissent in authoritarian regimes;[50] and the conditions that enable online disinformation campaigns to influence public opinion during social unrest.[51] By showing how counter-extremism activities encouraged IS sympathizers to behave strategically online, I contribute to a growing body of work that shows how conflict processes manifest in the online world. How militant groups use social media, and how efforts to prevent violence shape their activities in the online world, are likely to remain important research questions in the years to come.

Data Availability Statement

Replication files for this research note may be found at <https://doi.org/10.7910/DVN/TVMFQT>.

47. Altier, Thoroughgood, and Horgan 2014; Kimmel 2018; Mattsson and Johansson 2019.
48. Walter 2017; Zeitzoff 2017.
49. Pierskalla and Hollenbach 2013; Warren 2015; Weidmann 2015.
50. King, Pan, and Roberts 2013; Pan and Siegel 2020; Steinert-Threlkeld 2017.
51. Golovchenko, Hartmann, and Adler-Nissen 2018.

Supplementary Material

Supplementary material for this research note is available at <https://doi.org/10.1017/S0020818321000242>.

References

Altier, Mary Beth, Christian N. Thoroughgood, and John G. Horgan. 2014. Turning Away from Terrorism: Lessons from Psychology, Sociology, and Criminology. *Journal of Peace Research* 51 (5):647–61.

American Civil Liberties Union. 2016. What Is Wrong With the Government's "Countering Violent Extremism" Programs. <https://www.aclu.org/sites/default/files/field_document/cve_briefing_paper_-feb_2016.pdf>.

Berger, J.M. 2015. Tailored Online Interventions: The Islamic State's Recruitment Strategy. *CTC Sentinel* 8 (10):19–23.

Berger, J.M., and Heather Perez. 2016. Occasional Paper: The Islamic State's Diminishing Returns on Twitter: How Suspensions Are Limiting the Social Networks of English-Speaking ISIS Supporters. Program on Extremism at George Washington University, Washington, DC.

Bjelopera, Jerome P. 2014. Countering Violent Extremism in the United States. Congressional Research Service, Library of Congress. <https://fas.org/sgp/crs/homesec/R42553.pdf>.

Bloom, Mia, Hicham Tiflati, and John Horgan. 2019. Navigating ISIS's Preferred Platform: Telegram. *Terrorism and Political Violence* 31 (6):1242–54.

Briggs, Rachel. 2010. Community Engagement for Counterterrorism: Lessons from the United Kingdom. *International Affairs* 86 (4):971–81.

Carter, Joseph A., Shiraz Maher, and Peter R. Neumann. 2014. #Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks. International Centre for the Study of Radicalisation and Political Violence.

Council on American-Islamic Relations. 2016. Brief on Countering Violent Extremism (CVE). <https://www.cair.com/government_affairs/brief-on-countering-violent-extremism-cve/>.

Dalgaard-Nielsen, Anja, and Patrick Schack. 2016. Community Resilience to Militant Islamism: Who and What? An Explorative Study of Resilience in Three Danish Communities. *Democracy and Security* 12 (4):309–27.

Davey, Jacob, Jonathan Birdwell, and Rebecca Skellett. 2018. Counter Conversations: A Model for Direct Engagement with Individuals Showing Signs of Radicalisation Online. Institute for Strategic Dialogue.

Department of Homeland Security. 2015. Community Awareness Briefing: Foreign Fighter Focus. <https://www.aclu.org/sites/default/files/field_document/Community-Awareness-Briefing.pdf>.

Dunn, Kevin Mark, Rosalie Atie, Michael Kennedy, Jan A. Ali, John O'Reilly, and Lindsay Rogerson. 2016. Can You Use Community Policing for Counter Terrorism? Evidence from NSW, Australia. *Police Practice and Research* 17 (3):196–211.

Fernandez, Alberto M. 2015. Here to Stay and Growing: Combating ISIS Propaganda Networks. US-Islamic World Forum Papers 2015, Brookings Project on US Relations with the Islamic World.

Fishman, Brian. 2019. Crossroads: Counter-Terrorism and the Internet. *Texas National Security Review* 2 (2).

Gillum, Rachel M. 2018. *Muslims in a Post-9/11 America: A Survey of Attitudes and Beliefs and Their Implications for US National Security Policy*. University of Michigan Press.

Golovchenko, Yevgeniy, Mareike Hartmann, and Rebecca Adler-Nissen. 2018. State, Media and Civil Society in the Information Warfare over Ukraine: Citizen Curators of Digital Disinformation. *International Affairs* 94 (5):975–94.

Greenberg, Karen J. 2016. Case by Case: ISIS Prosecutions in the United States. Center on National Security, Fordham University School of Law.

Hamm, Mark S., Ramon F.J. Spaaij, and Simon Cottee. 2017. *The Age of Lone Wolf Terrorism*. Columbia University Press.
Helmus, Todd C., and Kurt Klein. 2018. Assessing Outcomes of Online Campaigns Countering Violent Extremism: A Case Study of the Redirect Method. RAND Corporation.
Jackson, Brian A., Ashley L. Rhoades, Jordan R. Reimer, Natasha Lander, Katherine Costello, and Sina Beaghley. 2019. *Practical Terrorism Prevention*. RAND Corporation.
Jurgens, David, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. *Proceedings of the Ninth International AAAI Conference on Weblogs and Social Media.*
Kimmel, Michael S. 2018. *Healing from Hate: How Young Men Get Into—and Out of—Violent Extremism*. University of California Press.
King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review* 107 (2):326–43.
Kundnani, Arun. 2009. Spooked! How Not to Prevent Violent Extremism. Institute of Race Relations. <https://www.kundnani.org/wp-content/uploads/spooked.pdf>.
Leetaru, Kalev, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, and Eric Shook. 2013. Mapping the Global Twitter Heartbeat: The Geography of Twitter. *First Monday* 18 (5).
Marwick, Alice, and Rebecca Lewis. 2017. Media Manipulation and Disinformation Online. Data and Society Research Institute, New York.
Mattsson, Christer, and Thomas Johansson. 2019. Leaving Hate Behind: Neo-Nazis, Significant Others and Disengagement. *Journal for Deradicalization* 18:185–216.
Mitts, Tamar. 2019. From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West. *American Political Science Review* 113 (1):173–94.
Nouri, Lella, Nuria Lorenzo-Dus, and Amy-Louise Watkin. 2019. Following the Whack-a-Mole: Britain First's Visual Strategy from Facebook to Gab. Paper no. 4, Global Research Network on Terrorism and Technology.
Pan, Jennifer, and Alexandra Siegel. 2020. How Saudi Crackdowns Fail to Silence Online Dissent. *American Political Science Review* 114 (1):109–25.
Patel, Faiza, and Michael German. 2015. Countering Violent Extremism: Myths and Fact. Brennan Center for Justice at New York University School of Law. <http://tinyurl.com/y6bpfexa>.
Pierskalla, Jan H., and Florian M. Hollenbach. 2013. Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa. *American Political Science Review* 107 (2):207–24.
Romaniuk, Peter. 2015. Does CVE Work? Lessons Learned from the Global Effort to Counter Violent Extremism. Global Center on Cooperative Security. <https://www.globalcenter.org/wp-content/uploads/2015/09/Does-CVE-Work_2015.pdf>.
Steinert-Threlkeld, Zachary C. 2017. Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring. *American Political Science Review* 111 (2):379–403.
Tech Transparency Project. 2020. White Supremacist Groups Are Thriving on Facebook, 21 May. <https://www.techtransparencyproject.org/articles/white-supremacist-groups-are-thriving-on-facebook>.
Thomas, Paul. 2010. Failed and Friendless: The UK's "Preventing Violent Extremism" Programme. *British Journal of Politics and International Relations* 12 (3):442–58.
Vermeulen, Floris. 2014. Suspect Communities: Targeting Violent Extremism at the Local Level. Policies of Engagement in Amsterdam, Berlin, and London. *Terrorism and Political Violence* 26 (2):286–306.
Vidino, Lorenzo, and Seamus Hughes. 2015. ISIS in America: From Retweets to Raqqa. Program on Extremism, George Washington University, Washington, DC. <http://tinyurl.com/y49ucbgv>.
Walter, Barbara F. 2017. The New New Civil Wars. *Annual Review of Political Science* 20:469–86.
Warren, T. Camber. 2015. Explosive Connections? Mass Media, Social Media, and the Geography of Collective Violence in African States. *Journal of Peace Research* 52 (3):297–311.
Weidmann, Nils B. 2015. Communication Networks and the Transnational Spread of Ethnic Conflict. *Journal of Peace Research* 52 (3):285–96.
Yayla, Ahmet S. 2017. The Reina Nightclub Attack and the Islamic State Threat to Turkey. *CTC Sentinel* 10 (3).

Zeitzoff, Thomas. 2017. How Social Media Is Changing Conflict. *Journal of Conflict Resolution* 61 (9): 1970–91.

Author

Tamar Mitts is Assistant Professor of International and Public Affairs at Columbia University. She can be reached at tm2630@columbia.edu.

Acknowledgements

I thank Kelly Berkell, Christopher Blattman, Lindsay Dolan, Page Fortna, John Huber, Macartan Humphreys, Egor Lazarev, Megan Metzger, Joshua Mitts, Suresh Naidu, Richard Nielsen, Arie Perliger, Kiki Pop-Eleches, Peter Romaniuk, Jacob Shapiro, Kunaal Sharma, Alexandra Siegel, Tara Slough, Jack Snyder, Zachary Steinert-Threlkeld, and Thomas Zeitzoff for their helpful feedback on the paper.

Key Words

Terrorism; violent extremism; social media; Islamic State; text as data

Date received: January 9, 2020; Date accepted: January 5, 2021