TECHNOLOGY FOR GLOBAL SECURITY

THE ONLINE TARGETING OF JOURNALISTS WITH ANTI-SEMITIC INTIMIDATION

Brittan Heller

TECHNOLOGY FOR GLOBAL SECURITY SPECIAL REPORT

OCTOBER 25, 2018

RECOMMENDED CITATION

Brittan Heller, "THE ONLINE TARGETING OF JOURNALISTS WITH ANTI-SEMITIC INTIMIDATION", T4GS Special Reports, October 25, 2018, http://www.tech4gs.org/social-media-storms-and-nuclear-early-warning-systems-114772.html

THE ONLINE TARGETING OF JOURNALISTS WITH ANTI-SEMITIC INTIMIDATION

BRITTAN HELLER, OCTOBER 25, 2018

I. Introduction

In this essay, Brittan Heller argues that an examination of online anti-Semitic attacks of journalists on Twitter, leading up to the 2016 U.S. presidential election, was one of the first indicators of the deliberate targeting of minority groups on social media. Examining this incident, with the benefit of hindsight, provides insights into the nature, purpose, and intended impact of online troll storms, including: professionalized trolling, enmity toward the professional press, "useful idiot"-based virality, and bridging online conduct and offline harms. Heller asserts that the incident should be instructive to decision-makers who aim to stem the tide of real-world violence, showing us what we can learn from the systematic online targeting of minority populations.

Brittan Heller is the director of technology and society for the Anti-Defamation League. She is an affiliate of the Berkman Klein Center for Internet and Society at Harvard University.

This paper was presented on October 20, 2018 to the *Social Media Storms and Nuclear Early Warning Systems Workshop* held at the Hewlett Foundation campus. The workshop was co-sponsored by Technology for Global Security, the Nautilus Institute, the Preventive Defense Project—Stanford University, and was funded by the MacArthur Foundation. This is the first of a series of papers from this workshop. It is published simultaneously here by the Nautilus Institute. Readers may wish to also read REDUCING THE RISK THAT SOCIAL MEDIA STORMS TRIGGER NUCLEAR WAR: ISSUES AND ANTIDOTES and Three Tweets to Midnight: Nuclear Crisis Stability and the Information Ecosystem published by Stanley Foundation.

II. T4GS SPECIAL REPORT BY BRITTAN HELLER: THE ONLINE TARGETING OF JOURNALISTS WITH ANTI-SEMITIC INTIMIDATION

OCTOBER 25, 2018

1. Summary

This is the story of a number — 2.6 million — and how it became one of the first indicators of the systematic targeting of minority groups on social media. Examining this incident, with the benefit of hindsight, allows us to draw several insights into the nature, purpose, and intended impact of online troll storms:

- Our examination was one of the early indicators that trolling techniques were being professionalized with the intent to intimidate and discriminate against minority groups. This has emerged as a primary tactic of online propaganda campaigns utilizing social media as their medium of choice.
- Early indicators of enmity toward the professional press, which also deepens as a part of social-media based information operations.
- Dangers of virality, regardless of the sourcing, when fueled by the "useful idiot" problem.
- Social media-based enmity was a bridge between offline conduct and online violence, which was demonstrated by the fatalities in Charlottesville a year later.

What follows will be a description of this attack and the response from the Anti-Defamation League (ADL), who conducting the first big-data based analysis of potential civil rights abuses. This will be followed by a discussion of the significance of such an incident and what advocates who mean to stem the tide of real-word violence can learn from systematic online targeting of minority populations.

2. The Attacks

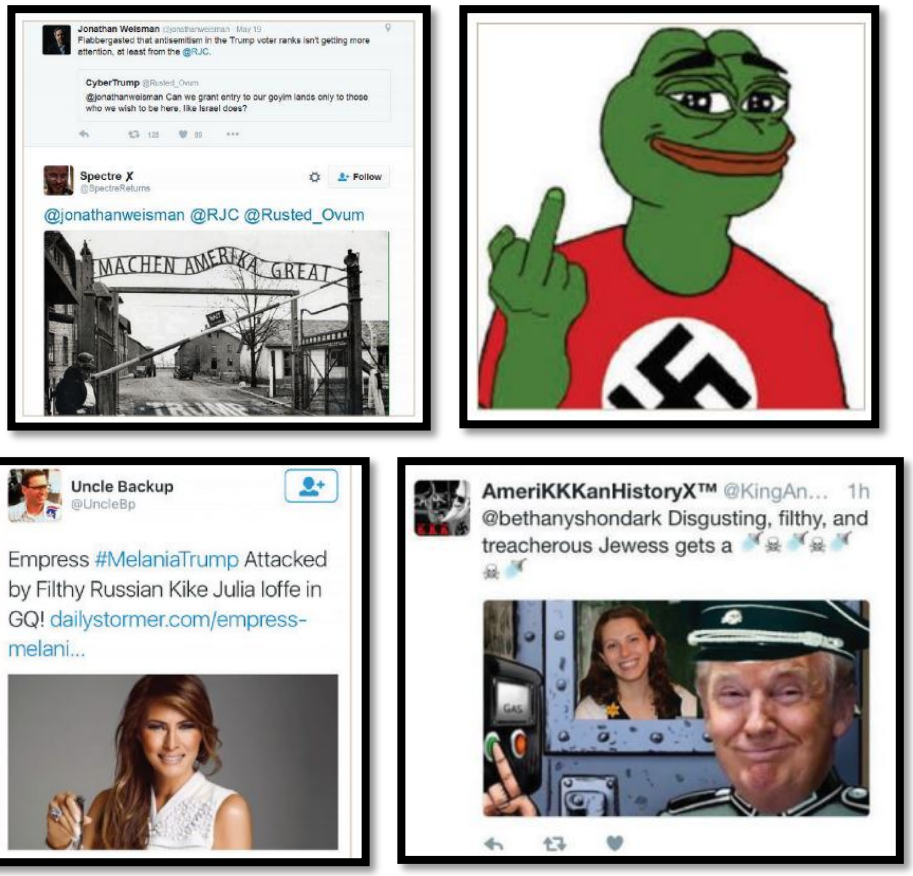
During the 2016 electoral season, it became dangerous to be a journalist. ADL received a rash of complaints from journalists, who were targeted with virulent messages on Twitter in the course of covering the campaign trail in the course of their work.[1] In particular, this impacted those journalists covering the presidential election. *New York Times* reporter Jonathan Weisman received images of himself in concentration camp ovens, with then-candidate Trump about to push the "on" button. He also received doctored images of himself wearing Nazi "Juden" stars, and of Auschwitz's infamous entry gates, the path painted over with the Trump logo, and the iron letters refashioned to read "Machen Amerika Great."[2]

The tweets crossed the line between online critique and offline abuse in a tangible and often frightening way. Julia Ioffe, after writing a profile of Melania Trump for GQ Magazine, found the abuse moved offline, receiving calls from coffin makers asking what size the child's coffin for her family order should be.[3] Kirk Eichenwald, a journalist who was openly epileptic, was sent videos featuring Pepe the Frog — a cartoon that had become the unofficial mascot of the alt-Right and was declared a hate symbol by the Anti-Defamation League.[4] These auto-playing videos contained flashing images designed to trigger Eichenwald's condition and, as a result, he suffered from seizures and was incapacitated for several months.[5]

It was not merely the anti-Semitic content of the abuse, but frequency of the Tweets, that was troubling. Independent journalist Bethany Mandel experienced 19 hours of continual abuse, and then went out and bought herself a gun.[6]

---

[1] *Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign,*
[2] *Control-Alt-Delete: Recommendations of the ADL Task Force on the Online Harassment of Journalists, https://www.adl.org/sites/default/files/documents/assets/pdf/press-center/adl-journalism-task-force-recommendations.pdf, 31* (December 2016).
[3] *Control-Alt-Delete,* 31 (2016).
[4] *How Donald Trump Supporters Attack Journalists,* https://www.newsweek.com/epileptogenic-pepe-video-507417, Newsweek (October 2016).
[5] *Control-Alt-Delete, 1* (2016).
[6] *Control-Alt-Delete, 36* (2016).



Photos Courtesy of Control-Alt-Delete: Recommendations of the ADL Task Force on the Online Harassment of Journalists

3. The Analysis

Self-reporting by the journalists strongly indicated that the rhetoric in the 2016 campaign had unleashed a torrent of hate and harassment against Jewish journalists online.[7] To examine this further, analysts at ADL decided to conduct a data-based analysis to determine the full scope of the problem; who specifically what being impacted; why this group was being targeted; where was the abuse coming from; when were the attacks occurring; why was this attack being implemented; and moreover, given the answers we received, how could ADL best address the situation?

Since the reports that ADL received had centered on Twitter-centric abuse, the examination began with a data pull of all tweets from August 2015 to July 2016 using a customized version of a commercially-available social media listening tool. To get at the underlying content, ADL compared this large corpus against a data set of anti-Semitic language, as determined by their subject matter experts.[8] In total, this

[7] *The Tide of Hate Directed Against Jewish Journalists*, https://www.theatlantic.com/politics/archive/2016/10/what-its-like-to-be-a-jewish-journalist-in-the-age-of-trump/504635/, The Atlantic (October 2016).

[8] Dictionary-based searches likely would not be the preferred methodology today, given the evolution of AI and machine learning techniques for studying online speech. Such a technique is limited in scope, as it does not account for shifts in meaning and context in hateful speech. However, at the time of the study, this technique allowed researchers to pare down the data set with a given set of known variables.

resulted in 2.6 million anti-Semitic tweets found in over a one-year period. The tweets had a reach of 10 billion impressions, meaning they came up on users screens at least that many times.[9]

Photo Courtesy of Control-Alt-Delete: Recommendations of the ADL Task Force on the Online Harassment of Journalists

This word cloud represents the frequency of use of certain terms by font size, as found in the analysis of the 2.6 million tweets. The most-used terms are depicted in a larger font. Researchers found that the type of language that is most prevalent in the examined tweets was demonstrably anti-Semitic in character.[10] Next researchers narrowed the analysis on how much of this anti-Semitic content was directed at known journalists. ADL took a roster of 50,000 known journalists and compared it to the 2.6 million anti-Semitic tweets.[11] This resulted in 19,253 overtly anti-Semitic hits targeting journalists. Researchers manually reviewed these hits and found that the hits were directed toward approximately 800 journalists. This figure is likely smaller than the actual frequency of abuse, since ADL erred on the side of a conservative analysis, and the methodology did not catch "coded language" that substitutes for anti-Semitic slurs or indicators of targeting.[12]

[9] Impressions are the preferred metric, as opposed to views, since ADL could not accurately represent that users saw these images, even if they occurred in the Twitter feed.
[10] *Control-Alt-Delete,* 33 (2016).
[11] *Control-Alt-Delete,* 33 (2016).
[12] *Ibid.*

Photo Courtesy of Control-Alt-Delete: Recommendations of the ADL Task Force on the Online Harassment of Journalists

Here is a word cloud of the top words used in the abuse of the journalists, again showing the most-used terms with the larger font indicating a higher frequency of use.

The results illuminated who the targets were and that their perceived Jewishness motivated the attack. Approximately 83% of the abusive tweets were targeting only ten journalists. These top ten targets were both men and women, from all parts of the political spectrum, who wrote for a variety of types of media publications.[13] The connection between them was the perception of Jewishness, either from their surnames, last names, or from their openness about their Jewish identity.[14]

[13] Ibid.
[14] Ibid.

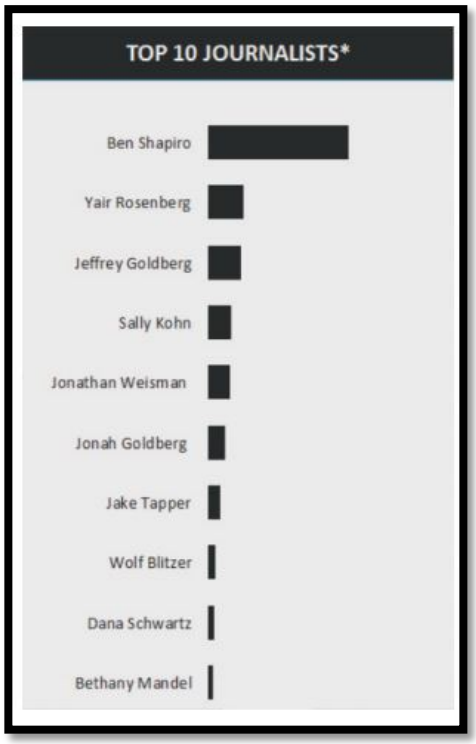


Photo Courtesy of Control-Alt-Delete: Recommendations of the ADL Task Force on the Online Harassment of Journalists

Researchers found a definitive link between offline events and increases in online abuse. The data shows that the upsurge in anti-Semitic hate speech and harassment was driven by the rhetoric in the presidential campaign.[15] ADL noted there was no known causal relationship between Mr. Trump or his campaign and the wave of anti-Semitic attacks against journalists. However, these self-appointed Trump surrogates used events in the campaign, especially actions by Mr. Trump, as a justification for attacking journalists. [16]

[15] *US Election Causing Uptick in Twitter Abuse for Jews and Journalists—Study,* https://www.theguardian.com/us-news/2016/oct/19/twitter-abuse-antisemitism-trump-alt-right-anti-defamation-league, The Guardian (October 2016).
[16] *Jewish Reporters Harassed by Trump's Anti-Semitic Reporters*, https://www.npr.org/2016/07/06/484987245/jewish-reporters-harassed-by-trumps-anti-semitic-supporters, National Public Radio (July 2016).

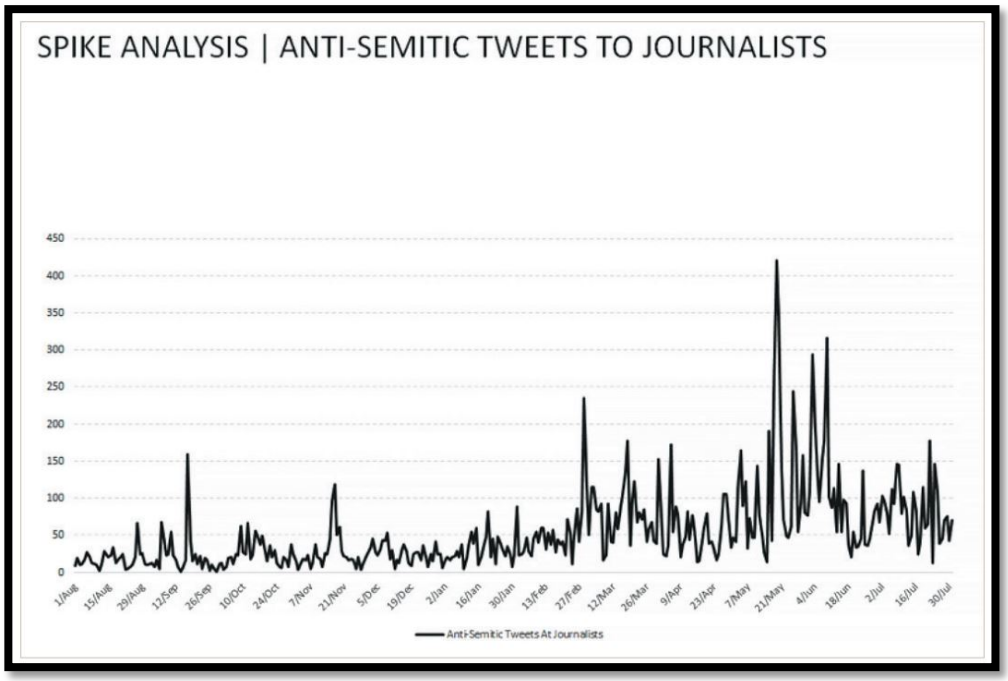


Photo Courtesy of Control-Alt-Delete: Recommendations of the ADL Task Force on the Online Harassment of Journalists

For example, this spike analysis indicates increases in the incidence of online abuse. There was a spike in anti-Semitic Twitter activity on February 29, 2016 during peak coverage of Trump's refusal to disavow the Ku Klux Klan.[17] Another spike occurred on May 17, 2016, when Melania Trump asserted that journalist Julia Ioffe "provoked" the anti-Semitic attacks against her.[18] A similar spike occurred on May 25, 2016, when Trump verbally attacked a federal judge whose parents emigrated from Mexico.[19]

Looking at the span of abusive tweets over time, there is a significant uptick in anti-Semitic tweets from January 2016 to July 2016 as the presidential campaign coverage intensified. Approximately 76% of tweets at journalists were posted in the February–July timeframe, as compared to the 24% of tweets at journalists posted between August and January. While the presence of anti-Semitic tweets spiked following election-related news events, the language used in the anti-Semitic tweets was not solely election-related. Many tweets referenced classic anti-Semitic tropes, like Jews controlling the media, manipulating control global finance, or perpetrated the attacks of 9/11.[20]

To determine the motivations of the attackers, researchers examined that population using their own words. Below see a word cloud created from the Twitter bios of unique users producing the abuse. Through this, shared characteristics in their profiles begin to emerge.[21] A majority of the harassment originates from accounts that identify themselves as Donald Trump supporters, conservatives, or part of the so-called "alt-

[17] *Donald Trump Refuses to Condemn KKK, Disavow David Duke Endorsement*, http://time.com/4240268/donald-trump-kkk-david-duke/, TIME (February 2016).
[18] *How Donald Trump Supporters Attack Journalists* (2016).
[19] *Trump's Attacks on Judge Curiel Are Still Jarring to Read,* https://www.cnn.com/2018/02/27/politics/judge-curiel-trump-border-wall/index.html, CNN (February 2018).
[20] *Control-Alt-Delete*, 35 (2016).
[21] *Don't Feed The Trolls? It's Not That Simple*, https://www.dailydot.com/via/phillips-dont-feed-trolls-antisocial-web/, The Daily Dot (June 2013).

right," a catch-all phrase that refers to white supremacists, racists and other extremists on the far right.[22] The words most frequently used in the attackers' Twitter bios are "Trump," "nationalist," "conservative," and "white." More information from the bios indicated that two-thirds of these abusers self-identified as male. [23]

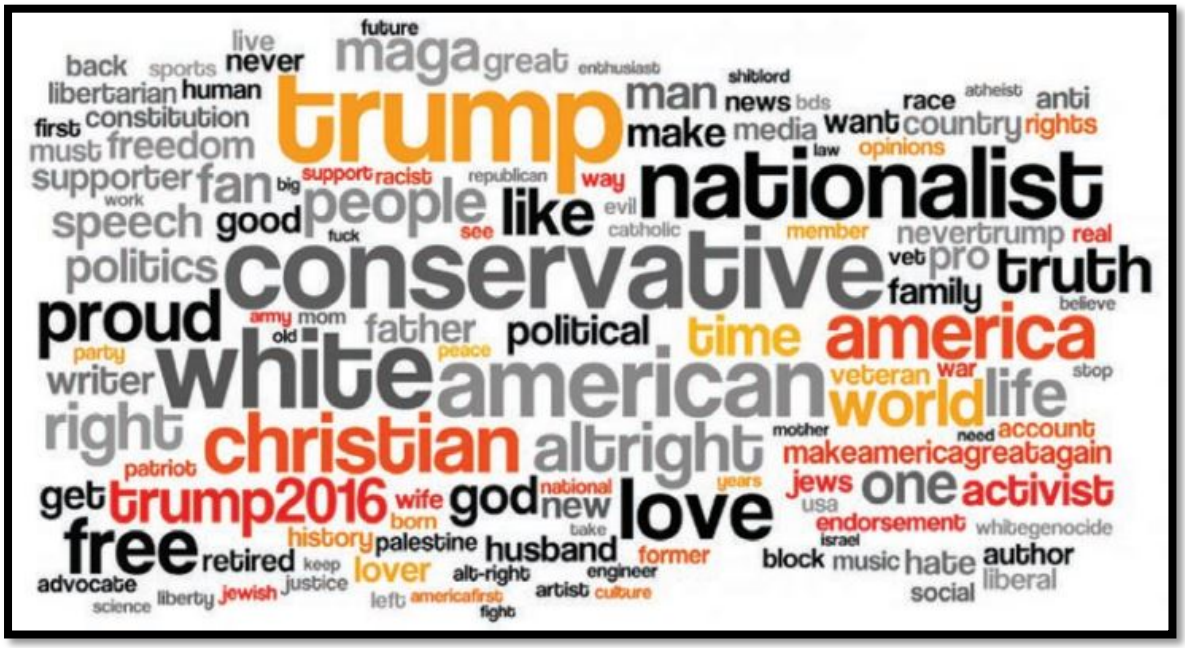


Photo Courtesy of Control-Alt-Delete: Recommendations of the ADL Task Force on the Online Harassment of Journalists

The presence of political slogans in the attacker's bios was an unexpected result. This indicated the influence of Donald Trump's use of Twitter as a campaign tool. Researchers were able to infer that attackers felt enabled to use Tweets as weapon against "enemies" of the campaign — and these "enemies" were, oftentimes, Jewish journalists.

Overall, the study found that a comparatively small group of attackers drove most of the anti-Semitic hate and harassment on Twitter, but these individuals had an outsized impact. More than two-thirds of the anti-Semitic tweets directed at journalists were sent by 1,631 Twitter accounts, out of 313 million total Twitter accounts at the time of the attack.[24] While this is a small proportion of Twitter users, the comparative impact of this abuse was widespread. The reach was tantamount to the spread covered by a $20 million dollar Superbowl ad.[25]

5. The Response

ADL took this analysis and used the data to tell the story of the harassment. ADL's Center for Technology and Society (CTS) was able to take these results directly to Twitter, in addition to other technology companies, journalists, tech policy advocates, government officials and law enforcement, in order to demonstrate that there was a discrimination-based attack that warranted a serious response.

---

[22] *The Alt-Right*, https://www.splcenter.org/fighting-hate/extremist-files/ideology/alt-right, The Southern Poverty Law Center.
[23] *Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign,* 8 (2016).
[24] *Control-Alt-Delete*, 29 (2016).
[25] *Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign*, 5 (2016).

Because CTS was able to ground our recommendations in empirical data, we were able to get a meaningful response from tech companies. Twitter responded to this report by meeting with CTS to engage with the behavior, and take preventative measures to curb future abuse for all its users, developing and introducing new tools to help users curb harassment and take control of their own feeds.[26] Del Harvey, Head of Trust and Safety, spoke to the Jewish community directly at ADL's national convention (focused on combatting anti-Semitism), explaining the new tools and the company's efforts to improve users' experience on Twitter.[27] ADL was also able to form an anti-Cyberhate Problem Solving Lab with both technical and legal expertise.[28]

Twitter also reviewed the information provided to them about the abusers. Because a small group of users was having a disproportionally large impact on a minority group — and arguably on civil society at-large due to the potential intimidation of journalists — Twitter felt empowered to take action. Twitter's staff reviewed the accounts producing the abuse and found even more than that of the ADL's research. In total, over 2000 accounts were found to violate Twitter's community standards, and were subsequently banned from the platform.[29]

CTS took the report and used it as a springboard, interviewing over 150 experts and releasing a follow-up set of recommendations for audiences from law enforcement, public policy, journalism, the tech industry, and victimized groups.

6. The Significance

From this seminal study, ADL became the first to report on the widespread targeting of Jewish journalists on social media. Seen retrospectively, with the benefit of hindsight, the attack was the beginning of the borrowing from the techniques commonly used in trolling. To "troll" means to make a deliberately offensive or provocative online post with the aim of upsetting someone or eliciting an angry response.[30] Trolling en masse has been done to silence voices and drive them away from public participation. Trolling also uses the targets' immutable characteristics — like racial, ethnic, religious, sexual orientation, sexual identity, and gender-based classifications — to scapegoat targets.[31] In essence, this attack was indicative of the professionalization of online harassment, as journalists were targeted for the first time in a coordinated, biased-based attack, with the aim of limiting the journalists' participation in the public sphere, with discriminatory overtones to the harassment.[32]

What we are seeing today is the next evolution in this phenomenon. The study was one of the first indicators that the targeting of journalists on social media had become both systematic and politically strategic. Throughout 2017-2018, it has become well-established that foreign information operations operated during

---

[26] *Announcing the Twitter Trust and Safety Council*, https://blog.twitter.com/official/en_us/a/2016/announcing-the-twitter-trust-safety-council.html (February 2016).
[27] Ibid.
[28] *Facebook, Google, Microsoft, Twitter and ADL Announce Lab to Engineer New Solutions to Stop Cyberhate*, https://www.adl.org/news/press-releases/facebook-google-microsoft-twitter-and-adl-announce-lab-to-engineer-new, The Anti-Defamation League (October 2017).
[29] *Twitter Today Starts Enforcing New Rules Around Violence and Hate*, https://techcrunch.com/2017/12/18/twitter-today-starts-enforcing-new-rules-around-violence-and-hate/, Tech Crunch (December 2017).
[30] *Troll Definition*, https://en.oxforddictionaries.com/definition/troll, English Oxford Living Dictionaries.
[31] *Don't Feed the Trolls? It's Not That Simple*, The Daily Dot (2013).
[32] *Control-Alt-Delete*, 36 (2016).

the 2016 election cycle.[33] Social media companies and intelligence sources identified substantial trolling activity emerging from Russia's Internet Research Agency, during this time period, with the goal of impacting U.S. elections.[34] The attacks exploited the architecture of online systems and the terms of services of the platforms.[35] Part of these trolls' tactics was to target minority groups, in order to amplify social cleavages and encourage civil unrest. Social media was a primary attack vector, especially content meant to go viral on Facebook and Twitter.[36]

A retrospective analysis of our data set against publicly-available repositories of Russian sources revealed minimal overlap. However, other ADL data sets focusing on online anti-Semitic activity did show substantive interaction with suspected foreign actors. Deleting the accounts may have stopped the hate against Jewish journalists, but it may have also lost evidence. The true impact may never be definitively known. Even Twitter itself may be unable to definitively attribute foreign trolling activity.[37]

With this information in mind, the anti-Semitic targeting of journalists on Twitter can be examined in a new light — one demonstrating the potential features of social media in instigating offline conflict, and illuminating what could have been done to mitigate the potential harm. Key lessons can be drawn from this episode, illuminating (a) what to expect in future attacks; and (b) what can be done about them. Overall, there are four characteristics of professionalized trolling:

(1) The source of the attack is difficult to attribute;
(2) The attacks are spread by scapegoating targets;
(3) The aim of the attack is to silence critics or perceived opponents;
(4) Social media architecture is used to target minorities.

In aggregate, these attributes combine to reinforce that there is a real connection between online targeting and offline harm.

Difficulties in attribution for trolling

The attack on journalists emphasizes the challenges in attribution of social media-based attacks. Attribution is more of an art than a science.[38] Facebook itself wrote the process of identifying threat actors is more complex than it may appear, based on common tactics in cyber exploits, and this is with all the proprietary information available on the back end to a private company about its users.[39] Alex Stamos elaborates:

---

[33] *How to Combat Fake News and Disinformation*, https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/, The Brookings Institute (December 2017).
[34] *Harmful Content: The Role of Internet Platform Companies In Fighting Terrorist Incitement and Politically Motivated Disinformation*, http://www.stern.nyu.edu/experience-stern/faculty-research/harmful-content-role-internet-platform-companies-fighting-terrorist-incitement-and-politically, NYU Stern Center for Business and Human Rights, 13 (November 2017).
[35] Ibid.
[36] *How Trolls Are Ruining The Internet*, http://time.com/magazine/us/4457098/august-29th-2016-vol-188-no-8-u-s/, TIME (August 2016).
[37] *Twitter Admits There Were More Than 50,000 Russian Bots Trying to Confuse American Voters Before the Election*, https://slate.com/technology/2018/01/twitter-admits-there-were-more-than-50-000-russian-bots-confusing-u-s-voters-in-2016.html, Slate Magazine (January 2018).
[38] *Facebook and the Attribution Conundrum, https://digitalguardian.com/blog/facebook-and-attribution-conundrum, Digital Guardian (August 2018).*
[39] *How Much Can Companies Know About Who's Behind Cyber Threats?*, https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/, Facebook (July 2018).

> The first challenge is figuring out the type of entity to which [Facebook is] attributing responsibility. This is harder than it might sound. It is standard for both traditional security attacks and information operations to be conducted using commercial infrastructure or computers belonging to innocent people that have been compromised. As a result, simple techniques like blaming the owner of an IP address that was used to register a malicious account usually aren't sufficient to accurately determine who's responsible.[40]

Facebook claims it looks at a variety of factors, including coordination between threat actors, "tool, techniques, and procedures," and technical forensics — but no one is dispositive in creating attribution for trolling.[41] Furthermore, once bad actors are made aware that they are being tracked, traced, or monitored, they tend to stop engaging in those behaviors or take further precautions to anonymize their online presence.[42]

Given these challenges from the inside perspective of platforms, the information available to third-party researchers provides just a glimpse through the keyhole. While this has benefits in protecting user privacy, user data can be anonymized and released to vetted third-parties.[43] Twitter and Facebook have just begun to do this, in a very limited fashion.[44] [45] However, initial collaborations have been fruitful for Facebook, who recently relied on outside experts from the Atlantic Council to identify 32 group pages and inauthentic accounts from Facebook and Instagram.[46]

For example, at the time of our study, we were unable to determine the geographic origins of the harassing tweets directed at journalists, due to the limited available information from Twitter. This analysis would have had to be done by social indicators, which we did not deem to be reliable.
Trolling attacks are also growing increasingly sophisticated, as more attention is paid to online attacks.[47] Use of VPNs and encryption-based masking may obfuscate the origins of attacks, even for the most sophisticated reviewers.[48]

The true danger with attacks attributable to foreign actors is that they may convince us that cyberhate is an outside problem. The most recent studies show that what may look like bot-based engagement is actually that of motivated speakers, echoing political messaging with speech and rapidity.[49] These are homegrown threats — not Russian bots.

---

[40] *Id*.
[41] *Removing Bad Actors on Facebook*, https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/, Facebook (July 2018).
[42] Ibid.
[43] *Facebook Ignites Debate Over Third-Party Access to User Data*, https://www.wsj.com/articles/facebook-ignites-debate-over-third-party-access-to-user-data-1521414746, The Wall Street Journal (March 2018).
[44] *Implications of Changes in Twitter's Development Policy*, https://gwu-libraries.github.io/sfm-ui/posts/2017-05-18-twitter-policy-change, The George Washington University's Social Feed Manager: Helping Researchers and Archivists Build Social Media Collections (May 2017).
[45] *Facebook Provided User Data to Some Third Parties After Promising to Limit Access: Report*, http://thehill.com/policy/technology/391533-facebook-provided-user-data-to-some-third-parties-after-promising-to-limit, The Hill (June 2018).
[46] *Removing Bad Actors on Facebook*, Facebook (2018).
[47] *The Future of Free Speech, Trolls, Anonymity, and Fake News Online*, http://www.pewinternet.org/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/, Pew Research Center (March 2017).
[48] Ibid.
[49] *Crackdown on Bots Sweeps Up People Who Tweet Often*, https://apnews.com/06efed5ede4d461fb2eac5b2c89e3c11, The Associated Press (August 2018).

What is most important, based on the confluence of issues around attribution and the fundamental ambiguity of online identity, is that a lack of attribution not be seen as a bar to action.
Closely-holding information makes studying the exploits of platforms and online systems difficult for the public, and has potential negative impacts on public trust, in light of the Cambridge Analytica scandal and electoral hacking.[50]

Targeted scapegoating of minority groups

Some of the targeted journalists experienced "doxxing" or "doxing," meaning the release of personal information to an online audience with the express or implicit suggestion that this information is weaponized for offline violence.[51] For example, one victim received a call to her home telephone number from a coffin maker, inquiring where the child-sized coffin that had been ordered in her family's name should be delivered.[52]
With hindsight, this anti-Semitic incident was one of the early indicators that online trolling techniques were being professionalized with the intent to intimidate online voices and discriminate against minority groups. Trolling techniques with deep roots in online harassment have emerged as a primary tactic of online propaganda campaigns utilizing social media as their medium of choice.[53] Racial, ethnic, and religious minorities have also found themselves at the receiving end of this targeting, just as we saw in this case.[54]

Trolling with the intent to silence
Regardless of the sourcing, social media companies should be aware of the dangers of virality, when fueled by the "useful idiot" problem.[55] Often times the source of trolling may not be what is most volatile about the activity. Rather, the inertia behind a trolling attack can come from the efforts of those who join in the activity — for expression of their beliefs, for outrage's sake, or simply for fun. These bandwagon joiners spread messages far and wide, and are thus known as "useful idiots."[56]
Some of the trolling techniques commonly applied to journalists or minority voices include: unleashing "tweetstorms" or "dogpiling," which is the marshalling of other Twitter users to target an individual en masse; "sealioning" of "JAQ-ing (just asking questions) off" or the overwhelming of individual online voices with a deluge of content, in a short period of time, and subsequently claiming the speaker is the one not engaging when she declines to respond; or "astroturfing," which is using automated or organic crowds of Twitter users to create the false impression of grassroots support for an idea.[57]
Furthermore, attacks on the press, in the time of an election, are meant to limit access to information by brute force. The study was one of the earliest data-based indicators of enmity toward the professional press, a tactic

[50] *Understanding the Facebook-Cambridge Analytica Story: Quick Take*, https://www.washingtonpost.com/business/understanding-the-facebook-cambridge-analytica-story-quicktake/2018/04/11/071f8c84-3d97-11e8-955b-7d2e19b79966_story.html?utm_term=.5d2a8566b3a8, The Washington Post (April 2018).
[51] *Control-Alt-Delete*, 18 (2016).
[52] *Journalist Who Profiled Melania Trump Hit With Barrage of Anti-Semitic Abuse*, https://www.theguardian.com/us-news/2016/apr/28/julia-ioffe-journalist-melania-trump-antisemitic-abuse, The Guardian (April 2016).
[53] *Control-Alt-Delete*, 6 (2016).
[54] *Women and Minorities as Targets of Attack Online*, https://www.nytimes.com/roomfordebate/2014/08/19/the-war-against-online-trolls/women-and-minorities-as-targets-of-attack-online, The New York Times (August 2014).
[55] *Putin's Useful Idiots*, https://www.washingtonpost.com/opinions/putins-useful-idiots/2018/02/20/c525a192-1677-11e8-b681-2d4d462a1921_story.html?utm_term=.c98c00e65ac1, The Washington Post (February 2018).
[56] Ibid.
[57] *How to Respond to Internet Trolling*, https://medium.com/activism-theories-of-change/how-to-respond-to-internet-rage-77a255f85793, Medium (August 2014).

that has progressively deepened as a part of social-media based information operations.[58] While targeting the press is increasingly common with governments around the world, it is especially nefarious when it uses proxy actors to further political aims the by trying to restrict the free flow of information and ultimately the ballot box.[59]

Misuse and abuse of internet architecture by bad actors

For bias-motivated trolling to be the most effective, it must manipulate the fundamental ways that platforms perform.

This study of Twitter attacks was a good example of how actors find weaknesses in online systems and exploit them to further their aims. Common tactics used by Twitter harassers include using simple misspellings to confuse automated content moderation-based filters; containing harassing material in memes and graphics, instead of easier-to-police text-based content; and reporting those who are actually the targets of abuse, when these individuals engage in counter speech or publicly engaging with the abuse that they have received.[60]

Furthermore, bots are becoming increasingly sophisticated and frequently challenging to identify as inauthentic actors on platforms.[61] Bot herders maintain distinct personalities and often experts cannot tell them from humans based on speech alone.[62]

7. Conclusions and Recommendations

After the study, we are already seeing results from the TF study and recommendations. Based on CTS's interactions with Twitter, the company has reported that harassment is down ten-fold, one year out from the time of the attack on journalists. However, some platforms are reticent to remove users, while others are resistant to any type of content moderation.[63] These types of environments attract bad actors, some explicitly, and so platforms should be aware of the ramifications of their policy choices.[64]

To prevent offline violence stemming from online activity, platforms need increased awareness of targeted abuse on social media, and should respond with increased ways for users to protect themselves.

Social media-based enmity was a bridge between offline conduct and online violence, which was demonstrated by the fatalities in Charlottesville a year later.[65] Whereas previously scholars would question the linkage between "dangerous speech" that catalyzes violence and freedom of expression, this tie is no longer considered to be mere speculation.[66]

Advocates like ADL need to bolster the public's understanding that white supremacy is the main kind of online extremism operating within the United States, and the one that is most often met with violent results. Extremists with white supremacist leanings tend to be savvy in the use of online platforms and unleashing

---

[58] *Control-Alt-Delete*, 21 (2016).
[59] *Trump Denies Russia Is Still Targeting US Elections*, http://nymag.com/daily/intelligencer/2018/07/trump-denies-russia-is-still-targeting-u-s-elections.html, New York Magazine (July 2018).
[60] *Control-Alt-Delete*, 13 (2016).
[61] *As Many As 48 Million Twitter Accounts Aren't People, Says Study*, https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html, CNBC (March 2017).
[62] *The Follower Factory*, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html, The New York Times (January 2018).
[63] *Social Media's Silent Filter*, https://www.theatlantic.com/technology/archive/2017/03/commercial-content-moderation/518796/, The Atlantic (March 2017).
[64] *Companies Finally Shine a Light into Content Moderation Practices*, https://cdt.org/blog/companies-finally-shine-a-light-into-content-moderation-practices/, The Center for Democracy and Technology (April 2018).
[65] *Charlottesville: One Year Later*, https://www.adl.org/resources/reports/charlottesville-one-year-later#introduction, The Anti-Defamation League (August 2018).
[66] *See, e.g., www.dangerousspeech.org.*

internet-based terror.[67] The Jewish journalists study is only one example of this type of behavior, and most experts still look to ISIS or Russian-affiliated groups as the form of extremism that society should be most concerned about — when data shows that home-grown threats should be treated seriously by tech platforms.[68]

As such, tech companies need to be collaborating on cross-platform issues, outside of counterterrorism and child sexual exploitation-related content, to the extent permitted by law. Additionally, new collaborations with civil society groups, like ADL's Center for Technology and Society, can be made possible by the mutual sharing of data and expertise, in a way that curbs attacks on minority populations and civic actors. Finally, platforms need to understand their role in the cycle of online to offline conduct. Speech leads to action, especially in echo chamber-like environment. Ultimately, as we saw in the targeting of Jewish journalists, words matter.

III. T4GS INVITES YOUR RESPONSE

Technology for Global Security invites your responses to this report. Please send responses to: info@tech4gs.org. Responses will be considered for redistribution to the network only if they include the author's name, affiliation, and explicit consent.

[67] *The Alt-Right Doesn't Need To Be Visible To Succeed*, https://www.wired.com/story/alt-right-doesnt-need-visible-to-succeed/, WIRED (August 2018).
[68] *Control-Alt-Delete*, 12 (2016).