Computational Propaganda and the 2020 U.S. Presidential Election: Antisemitic and Anti-Black Content on Facebook and Telegram

Chapters

EXECUTIVE SUMMARY

By

Mark Kumleben and Samuel Woolley

Propaganda Research Team

Center for Media Engagement

University of Texas at Austin

Maggie Engler

Global Disinformation Index

Much of online speech today occurs on social media platforms like Facebook where a few companies have attained an outsized influence on what is permitted discussion on the internet. The rules and enforcement of the rules around speech on social platforms have far-reaching societal implications -- they can determine how a group of people are discussed, represented, and depicted. And in turn, tech company enforcement of these rules, or lack of enforcement, can have broad influence over the perspectives that individuals glean about other groups through exposure to online discussion.

For this report, we wanted to explore the scope and nature of online discussion about Jewish and Black communities in America through the lens of how often hateful content denigrating Jews and Black people marks those conversations. We conducted our analysis on two very different social platforms -- public pages on Facebook and private channels on Telegram. Using search terms to select discussions about Black and Jewish people in the United States, we collected 3,317 posts from 11 Facebook pages and 44,244 posts from 23 public Facebook groups that included some hateful or controversial content. We also collected 333,325 messages from 52 Telegram channels. Data was collected for the time period beginning January 1, 2020 and ending September 1, 2020.

Key findings include:

- Of the Telegram messages surfaced by our list of search terms, 26 percent of the messages containing references to any of the terms relating to Black people in America also contained derogatory language towards Black people,

and 57 percent of the messages containing references to any of the terms related to Jewish Americans also contained derogatory language towards Jewish Americans.

- 1 in every 81 messages sent to the Telegram channels we tracked were derogatory towards Black people in America, and about 1 in every 54 messages were derogatory towards Jewish Americans.
- Of the Facebook page and public group posts surfaced by our list of search terms, 7 percent of the messages containing references to any of the terms related to Black people contained derogatory language, and 9 percent of the messages containing references to any of the terms related to Jewish Americans contained derogatory language towards Jewish Americans
- 1 in every 150 posts tracked were hateful towards Black people on Facebook, and more than 1 in every 300 posts were derogatory towards Jewish Americans.

We also developed a set of recommendations for social platforms to reduce hateful content and to provide resources needed by researchers to continue exploring the online hate ecosystem. These are:

- Offer more anonymized data. Such tools presumably exist internally to allow proper content moderation, and could be modified for external research consistent with ethical bounds.
- Facebook should reconsider its policy of simply deleting hateful pages in favor of a policy that provides some record for researchers. Some of the pages we identified were taken down in the few days between searching for and scraping the pages.
- Carefully consider the consequences of tools made to counter online hate speech being used by authoritarian states to surveil, censor and penalize dissidents or other vulnerable users.
- Ameliorating hate cannot be left to any one service. Tech companies should design their products to prevent hateful content multiplying inside defined

online spaces like Facebook groups as well as hate movements from 'breaking out' into public social media. In particular, since harassment can be coordinated on one service and executed on another, all social media companies should design their products to be more harassment-resistant.

Government and civil society, including researchers also have an important role to play in curbing toxic discourse around particular identity-based groups:

- Governments at the federal, state and local levels need to adopt more comprehensive deradicalization initiatives. These need to understand the role various forms of hate and hate-adjacent content play. They should be focused on preventing vulnerable people from joining extreme groups online, deradicalizing users, regardless of whether they are members of online hate groups, and to prevent radicalization to the point of violence. These initiatives need to be adequately resourced, so sufficient appropriations are essential.
- Key stakeholders, from policy makers to community leaders and other influencers, need to use their bully pulpits firmly, persuasively and consistently to debunk hate-adjacent conspiracy theories, and attempt to do so in ways that are most effective, including to those vulnerable to radicalization.
- Researchers, journalists and others with relevant expertise should map and monitor extremist groups and relevant government programs should formally include diverse civil society and vulnerable community participation and input. While it is ultimately up to the government and law enforcement agencies to monitor and prevent potential terrorists, mapping extremist groups allows us to understand how they introduce their messages into public discourse and how they recruit new members. Government should appropriate resources to underwrite training of government officials and law enforcement agencies by such researchers.
- Given the extraordinary risk posed to individuals, communities and democratic guardrails, Government must continue to play a role -- through

public hearings and other means -- with respect to public oversight and the transparency and accountability of tech companies, particularly with regard to the development and enforcement of cyberhate policies. Such a role, however, must be consistent with constitutional prohibitions against Government censorship and overreach.

INTRODUCTION

In addition to the usual chaos wrought by a presidential election year, 2020 has been one of the most tumultuous years in recent history, as tensions over racial justice and police brutality boiled over into protests across America. In the wake of the murder of George Floyd by Minneapolis police officers, the Black Lives Matter movement (BLM), which originated on social media in 2013 and continues to engage heavily in online activism, captured the nation's attention. However, the inevitable blowback from racist hate groups was quick to arrive. A few days after the May 25 murder of George Floyd, one Telegram user messaged, "[A]s cities are burning in Minneapolis from black rioters with mass lootings, lets [sic] remember that the FBI will make no rioting arrests but yet will make a big show of arresting 3 members of Rise Above Movement." Another user claimed that "Jewish financed and led antifa, BLM, and communists are tearing down a statue of Civil War General Albert Pike in Washington DC right now." The rapid spike in anti-Black messages dovetailed with hate groups' consistent antisemitic preoccupations.

This study aims to analyze posts in hate or hate-adjacent groups on Telegram and Facebook, looking at both their anti-Black and antisemitic content. We label such content as 'derogatory': by which we mean pejorative, slur-laden language specifically aimed at demeaning and attacking particular social, racial, religious, and ethnic groups. We define hate-adjacent groups as groups that have not been categorized as hate groups by ADL, the Southern Poverty Law Center (SPLC), or similar entities but have been consistently reported on (by newspapers of record and civil rights groups) as either having connections to known hate groups or that often share hateful content, but for which hate/racism is not a core-focus or

direct-focus. Facebook and Telegram are quite different platforms, with the former focused on easily visible communication via newsfeeds or pages, and the latter on private messaging. Thus, we can see how hateful and derogatory content appears both in general online discourse and on the darker fringes of the internet. By better understanding the variation between public and private platforms, and the distinctions between the baseline types and levels of online hate and that sparked by a controversial event, we may be able to better mitigate the harm caused by online hate content and groups. One key finding of this report is that hateful content comprises a much larger proportion of references to Jewish and Black people in America on Telegram channels than in public Facebook groups, but that in all datasets derogatory references to Black people rose sharply as BLM became a national news story in May 2020.

LITERATURE REVIEW

The ugly realities of antisemitism and anti-Black racism are far from new in the United States, but recent political events have brought both to the forefront. Even in 2015, before the rises in hate documented in recent years, 10 percent of American adults professed antisemitic beliefs, and 33 percent believed that Jews are more loyal to Israel than to the US - centuries-old tropes maligning Jewish-Americans in political life (ADL, 2015). Since the 2016 election, however, antisemitism has proliferated both offline and online, with the ADL recording an over 100 percent increase in antisemitic incidents between 2015 and 2019 (ADL, 2020). At 2,107 separate incidents, this represents the highest level of antisemitic incidents reported by ADL since it began tracking in 1979, continuing an upward trend that began in 2016. Neo-Nazi groups, in particular, have increased in prevalence, with their number rising from 99 to 121 between 2016 and 2017 (SPLC, 2018). Much of this occurs on social media platforms, with, for example, researchers finding 4.2 million antisemitic tweets posted between 2017 and 2018 by roughly three million unique accounts (ADL, 2018). Rising levels of hate do not limit themselves to online content, however. In 2019, the ADL recorded a 56 percent year-over-year increase in antisemitic assaults, resulting in 95 victims

and five deaths (ADL, 2020). Antisemitic incidents have previously bubbled up around elections - during the 2018 midterms, such incidents peaked in months which did not show significant spikes in 2017 or 2019 (ADL, 2020), meaning that the 2020 elections are in danger of producing another wave of hate.

While antisemitism has continued to fester in the darker corners of U.S. politics, often without sufficient public attention, concerns over race relations erupted in 2020, sparked by high-profile police killings of Black Americans such as George Floyd and Breonna Taylor. The Black Lives Matter movement - which began in reaction to the 2013 acquittal of Trayvon Martin's shooter, George Zimmerman, and gained major traction after the death of Michael Brown in Ferguson, Missouri in 2014 - has become the focal point of the fight for racial equity, with sharp partisan divides emerging over BLM protests (Parker, Horowitz, & Anderson, 2020).

BLM began as a Twitter hashtag, #blacklivesmatter, and social media activism was integral to its early growth and widespread adoption (Day, 2015). Indeed, in the case of BLM, social media was critical in "helping far-flung activists develop a sense of collective identity" (Mundt et al., 2018). However, discussions on social media also spawned counter-protest opinions on hashtags such as #alllivesmatter (Gallagher et al, 2018). The darker side of social media was also evident, with the hashtag #blacklivesmatter skillfully manipulated in attempts to sow divisiveness and fear by the Russian Internet Research Agency's influence operations in the 2016 elections (Arif et al., 2018). Online extremism can also cross into deadly hate. Dylann Roof, the shooter who murdered nine African-American churchgoers in Charleston, South Carolina in 2015, was a 'lone wolf' extremist radicalized entirely online (Potok, 2015). This and other mass murders are then discussed extensively in far-right spaces, both as inspirations and as sources of hateful humor, inspiring other extremists to commit terrorist acts. Often, as in the case of the Christchurch, New Zealand mass murder in two mosques, white supremacists explicitly cite previous attackers as inspiration (ADL, 2019).

While many social media companies that have anti-hate policies still have less than stellar records keeping hate off their platforms, hate goes completely unchecked on websites like the far-right Twitter-alternative Gab and 4chan's /pol/. Between July 2016 and January 2018, the use of the terms "Jew" and "kike" more than doubled on 4chan and significantly increased on Gab (Finkelstein et al., 2018). Major social media websites, while still falling severely short, have improved a number of their stated policies over time - for instance, Facebook has recently announced its decision to ban Holocaust denial from its platform (ADL, 2020). Nonetheless, policies are only as effective as their enforcement and, in any case, are rarely applied to services that may be hosted and operated by companies otherwise outside of the platforms' direct control. This includes both encrypted messaging services owned by social media companies, like WhatsApp (Facebook), and dedicated private messaging tools like Telegram. Researchers have called Telegram "a safe space for hate", finding, for example, that 60.1 percent of white supremacist channels studied had glorified terrorism (Guhl & Davey, 2020). It may be that hate doesn't disappear when kicked off big platforms, but goes underground in purpose-built alternatives or on security-oriented apps. This, then, can prove to be an unintentional consequence of security tools or anti-hate policies.

METHODOLOGY

This report is designed to analyze social media data from hateful or potentially hateful groups, in order to better understand racist and antisemitic content online. We collected posts from Facebook pages, public Facebook groups, and Telegram channels, subjecting them to similar analysis in order to compare the forms hateful content takes on each platform. Our goal was to produce comparable data sets quantifying the proportion of references to Black or Jewish Americans that are derogatory, tracking hate across platforms and across time.

For this report, we collected 3,317 posts from 11 Facebook pages and 44,244 posts from 23 public Facebook groups using Facebook's CrowdTangle tool. This tool

provides content from social media pages to identify trends, and has been used effectively by hate researchers (Chandaluri & Phadke, 2019). We also collected 333,325 messages from 52 Telegram channels using the open-source Pushshift Telegram Ingest API. All sources chosen were expected to contain at least some hateful content. Facebook pages and groups were chosen by searching for terms identified by the Global Disinformation Index's (GDI) algorithms as associated with hateful content, and Telegram channels were selected based on a list provided by Memetica, a digital investigations consultancy. These were then searched for terms which could be used to describe Black or Jewish Americans, including neutral descriptors, racial slurs, and prominent individuals, organizations or movements, such as "Soros" or "BLM". Our goal was to find all references possible within a source, so false positives (such as a reference to the black clothes of Hong Kong protestors) were included to be classified later. All posts from these pages between 01/01/2020 and 09/01/2020 were filtered using these terms to produce data sets for Facebook pages, public groups, and Telegram channels.

From these data sets, we produced two stratified random samples for each, one stratified by search term and one by month. Each post in the sample sets was then coded by our researchers as derogatory or non-derogatory, allowing us to estimate the percentage of posts containing hateful content for a given term, as well as the percentage of posts containing hateful content in a given month. Coding was based specifically on whether the post included a hateful reference to Black or Jewish people - posts relating to other groups or containing generically white nationalist messages were not included. Stratifying by term allows us to discover which terms are more likely to be used in derogatory messages, and stratifying by month shows us changes in the prevalence of hateful content over 2020. For each term and month subsample, we construct an estimate of the proportion of the references that contain hateful content from our sample.

DATA ANALYSIS

Overall volume

Facebook

Of the Facebook page and public group posts surfaced by our list of search terms, 7 percent of the messages containing references to any of the terms related to Black people contained derogatory language, 9 percent of the messages containing references to any of the terms related to Jewish Americans also contained derogatory language directed at Jewish Americans. The 95 percent confidence interval for the former proportion is 5 percent to 9 percent given our sample size and 7 percent to 11 percent for the latter. Over the 11 pages and 23 groups we collected, there were 6,771 references to any of the terms related to Black people of which we expect about 494 to be derogatory based on the proportion of derogatory references in the samples manually coded. There were 2,556 references to any of the terms related to Jewish Americans, so despite that the percentage of offensive posts was higher for this category, we expect a lower total volume of derogatory posts, about 242 posts across our sample set. It should be noted that the percentage of references that are derogatory is compared to all references using the selected terms, regardless of whether or not they are specifically speaking of Black or Jewish Americans. Some of the more multi-faceted words, like “Black,” had plenty of unrelated false positives such as the use of words like “blackmail.” To place this in the context of the total conversation volume, more than 1 in every 150 posts were hateful towards Black people, and more than 1 in every 300 posts were derogatory towards Jewish Americans. Given the relatively restrictive definition we have used in this research, each of the derogatory references may reasonably be considered hateful content.

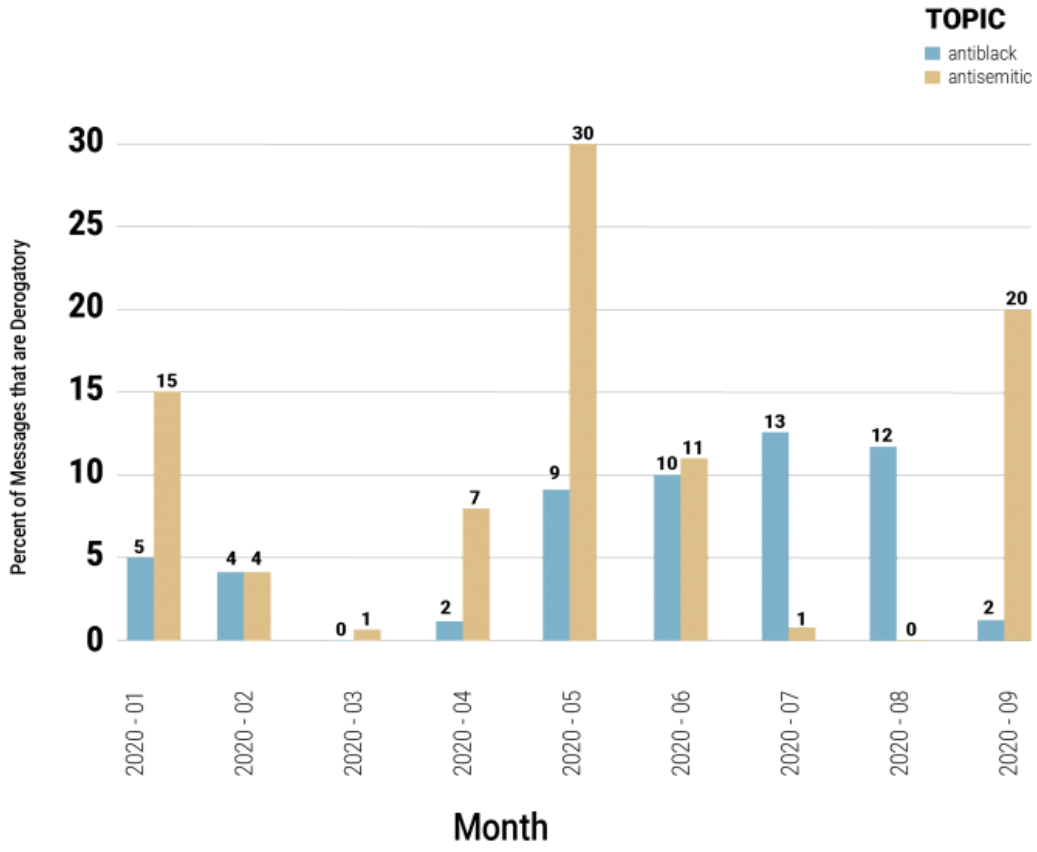
Telegram

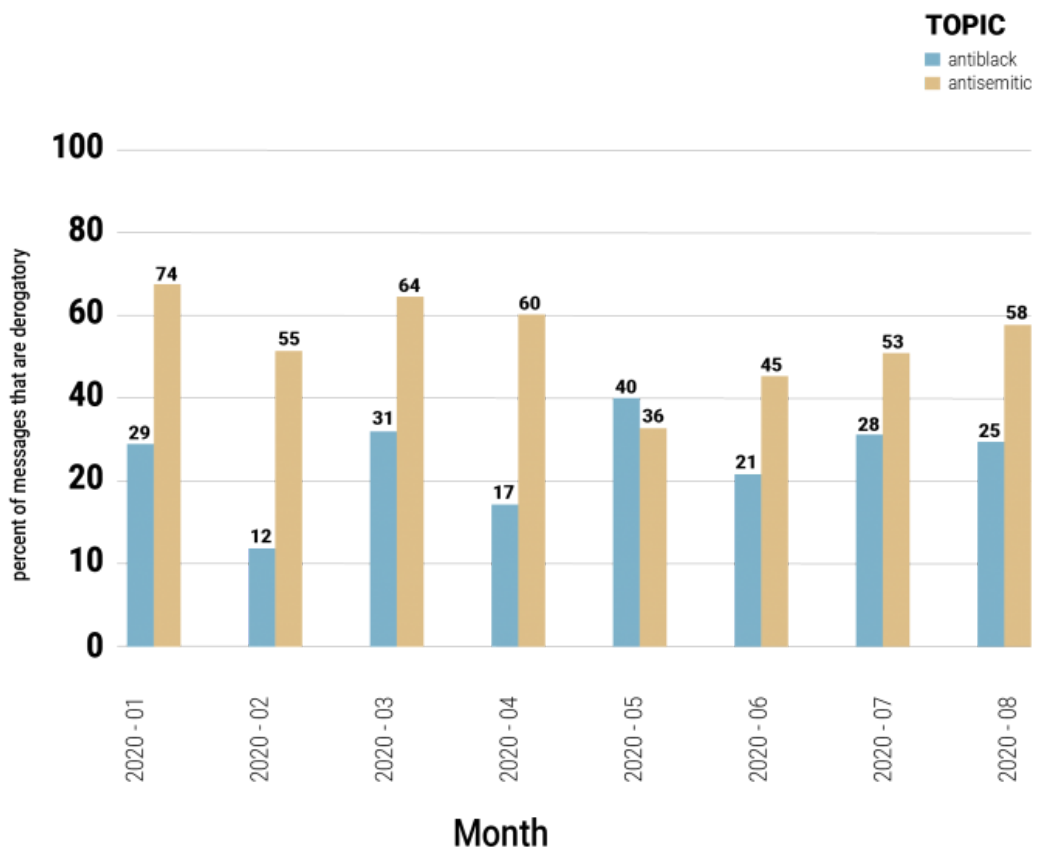
Of the Telegram messages surfaced by our list of search terms, 26 percent of the messages containing references to any of the terms related to Black people in America contained derogatory language towards Black people, and 57 percent of the messages containing references to any of the terms related to Jewish

Americans contained derogatory language towards Jewish Americans. The 95 percent confidence interval for the former proportion is 23 percent to 29 percent given our sample size and 53 percent to 61 percent for the latter. Over the 52 channels, there were 32,530 references to any of the terms related to Black people in American, of which we expect about 8,416 to be derogatory. There were 22,484 references to any of the terms related to Jewish Americans, of which we expect about 12,764 to be derogatory. Making the same comparison as above, about 1 in every 81 messages sent to the Telegram channels were derogatory towards Black people, and about 1 in every 54 messages were derogatory towards Jewish Americans. Additionally, for our Telegram collection in particular we were limited in our analysis by the fact that the collection mechanism did not include non-textual elements of posts, like memes and other images. If the message was not derogatory even with this additional context missing, it was not counted as such in our sample, and therefore this is likely an underestimate of the true proportion of derogatory conversations.

Change in content volume over time

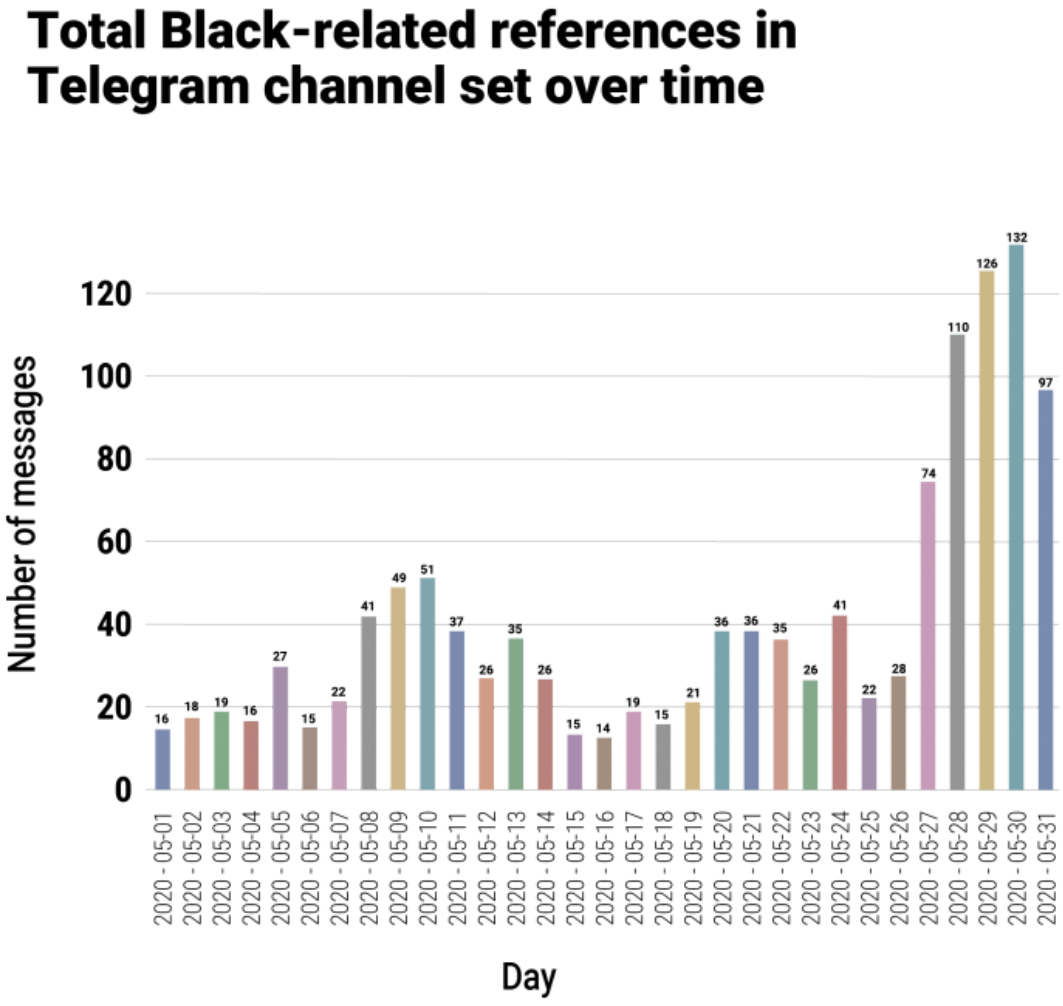
Derogatory references in Facebook set over time





The charts above show the number of derogatory references in our Facebook and Telegram collections over time. Note that the Telegram collections, with greater message volume and greater rates of derogatory messaging, maintain a much higher volume of derogatory content across the observed period. The most notable change across time in our data is a spike in derogatory anti-Black posts in June 2020 — which we think could be a response to the May 25th murder of George Floyd by Minneapolis police of George Floyd and subsequent Black Lives Matter protests. On Facebook pages, the number of derogatory posts in our sample quadrupled and stayed consistently at that level, from around 20 posts per month to around 80 (though we note that we have only three months of observation after this increase, and the months priofr to it show high variance despite the fairly low baseline). Telegram showed a similar increase in raw numbers of derogatory posts, at a larger scale, although the biggest jump in volume happens from April to

May 2020, then remains high through June, July and August. In fact, if we consider only the month of May 2020, we can pinpoint the response time even more precisely.



Antisemitic content did not show major changes over time, except for a substantial increase in derogatory posts on both Facebook and Telegram during January 2020. There is not an obvious explanation for this outlier month, though January did see the impeachment trial of President Donald Trump in the Senate as its major political event. Both Senators Chuck Schumer, the Senate Minority Leader, and Dianne Feinstein, the Ranking Member of the Senate Judiciary Committee, are Jewish, and their prominence in the news that month may have inflamed antisemitic conspiracy theorists.

Differences in term usage

Different terms referring to the two groups in question had very different likelihoods of containing derogatory content. Certain dogwhistle terms, such as “dindu” (referring to a racist meme common after police shootings) and “jogger” (referring to the February 2020 killing of Ahmaud Arbery), were highly likely to indicate derogatory posts - in the Telegram data, every single mention of “dindu” was derogatory. And to illustrate the extent to which “jogger” became a racial slur in this Telegram sample, the same percentage of posts referencing the words “jogger” and “negro”, were classified by our labelers as derogatory (80%). The only other similarly telling correlation was “Soros”, but only in the Facebook group data. George Soros, a Jewish Hungarian-American financier and philanthropist who is closely associated with progressive causes and whose Open Society Foundations funds civil society groups around the world, has been vilified by conspiracy theorists as central to the ‘undermining’ of western countries. These conspiracy theories, including the accusation that Soros is paying BLM and other “far left” protesters to riot, are reminiscent of Nazi propaganda (ADL, 2020).

On Telegram, in contrast, the term most likely to be used by an antisemitic post was simply “jew”. This might be explained by considering the apparent difference between hateful content on each platform. On public Facebook, groups are more likely to endorse conspiracy theories which may invoke antisemitic tropes, but are less likely to express Jew hatred in as coarsely direct a manner as they would in private groups. Fringe Telegram groups, however, are often more likely to directly denigrate Jews in ways Facebook’s content moderation standards do not permit.

CONCLUSION AND RECOMMENDATIONS

Our results underscore the fact that online hate -- particularly that which targets vulnerable communities that have long been subjected to virulent bigotry, such as Jews and Black people -- remains an issue worthy of concern. This study focused

on anti-Black racism and antisemitism, but similar reviews of Islamophobia, anti-LGBTQ+ content, anti-immigrant hate and misogyny, are also warranted.

Our report finds that a significant portion of online conversations about Jewish and Black people in America in the groups we tracked continue to be marked by hateful and derogatory content on Facebook and Telegram. This poses challenges to the ways that the Jewish and Black communities (and presumably other minority groups) are depicted, represented and discussed on the internet. Despite Facebook's repeated assurances of increased content moderation efforts around the election, the level of denigration and hate on both its platform and on Telegram leaves a lot to be desired. This is all the more alarming in the case of Facebook, as that data was collected from public groups that are presumably subject to content moderation, unlike private channels on Telegram which appear to be completely unmoderated. At the same time, because social networks create both organic and algorithmic silos and echo chambers, the typical experience of most users on social media might be one less likely to be marred by toxic content of the sort we found, especially to users who are not themselves part of a marginalized or vulnerable group identity.

However, it is clear that hate flourishes on both public and private spaces on mainstream social platforms like Facebook, or can be exiled to fringe platforms like Telegram. Additionally, major developments like the 2020 Black Lives Matter protests set off venomous reactions on the part of hatemongers and followers. Indeed, it is important consider how millions of American users who rely on Facebook for their daily intake of news are impacted if 7 percent of the messages in controversial or hateful groups containing references to any of the terms related to Black people are characterized by derogatory language, and 9 percent of the messages containing references to any of the terms related to Jewish Americans included derogatory language towards Jewish Americans. Similarly, if an alarming 26 percent of content about Black people contains anti-Black language and an extraordinary 57 percent of content about Jewish people contains antisemitic language on Telegram, how does that impact the world views

of users who rely on this platform to inform their perspective? This initial research doesn't provide answers to these questions, but it emphasizes how important it is to ask them and identify the impact of these pervasive networks.

In contemplating these questions, as well as the ways to disrupt, counter, remove or otherwise mitigate the effects of hate online, it is also important to understand the difference between the platforms studied (Facebook and Telegram). The Facebook pages and groups we analyzed were open to the public on a popular platform, whereas the Telegram channels we reviewed operate in obscurity (often intentionally). The difference between public-facing and isolated social networks as compared to private or secret groups, is an important one when considering their potential harmful effects of hateful content. Hateful content in public areas of the internet is far more widely accessible to all users of those websites, distressing some, potentially radicalizing others, and, overall, contributing to the normalization of hate and bigotry. Significantly, it is much easier to use sites like Facebook and Twitter to conduct harassment campaigns against members of minority groups, particularly public figures such as journalists, because of the public nature of the content on these platforms (ADL, 2018). Harassment campaigns can do real damage to the democratic process, silencing already marginalized voices and views at critical times.

Private, inward-looking groups (like those on Telegram), on the other hand, may not harm a platform's "ordinary" users, but they can pose a risk of far greater radicalization and extremist mobilization -- in large part because the content shared in these private groups is hidden from public view and potential opprobrium or countermeasures. As we have repeatedly seen, extremist content, often coupled with praise of terrorists and mass killers that is shared among groups of extremists connected to each other only by the internet, poses a risk of inciting lone wolf attackers or copycats inspired by previous shootings. Thus, the proliferation of hate on each platform has different effects and poses different dangers.

Recommendations for Technology Companies

While Facebook is finally taking more steps to actively moderate hate content, it is also keeping potentially important data out of the hands of researchers. Facebook's CrowdTangle, a content monitoring tool, neither includes individual user comments on posts, nor posts from private groups or from most Instagram accounts, deficiencies for which it has been criticized by researchers (Wagner, 2020). It may well be that individual users, less concerned about moderation than page owners, are posting hate in the comments of posts where we did not find the posts derogatory -- meaning that we did not then proceed to review comments.

Private groups are also extremely important parts of the hate ecosystem on Facebook because they have less oversight and create deep echo chambers. Most large groups found in our search for hate-adjacent terms were private. In fact, even non-controversial large groups are often made private due to fears that they could be subjected to false "mass-reporting" for hate speech, leading to automatic bans - a practice pioneered by large humor groups after a 2019 incident of mass-reporting by Indonesian trolls (Koebler & Haskins, 2019). There is a fine line to walk between protecting the privacy of individual Facebook members and more aggressively enforcing policies across a platform's services and allowing access for academic research. Facebook could offer at least some anonymized data. Such tools presumably exist internally to allow proper content moderation, and could be modified for external research within appropriate ethical bounds.

Finally, Facebook should reconsider its policy of simply deleting hateful pages in favor of a policy that provides some record for researchers. Some of the pages we identified were taken down in the few days between searching for and scraping the pages - indeed, if the paucity of data on hateful pages is due to rapid moderation by Facebook, the company should welcome researchers analyzing its success. On the other hand, if hate still flourishes in comment sections and private groups, it is important that it not be allowed to fly under the radar.

Combating hate in more private groups and privacy-oriented apps is a more difficult task than moderating public pages. There are benefits and drawbacks to pushing for Telegram to moderate content, as this could create unexpected harms. For instance, despite the fact that our study used a list intended to find American hate groups, it also captured a number of messages which appeared to be from Hong Kong groups offering support or advice to anti-CCP protestors. Any attempt to unmask extremists must be very careful not to endanger dissidents abroad by unintentionally enabling authoritarian governments to identify and persecute them. Telegram has recently stepped up efforts to identify and remove terrorist content (Europol, 2019), but it has also pledged to help the Russian government do likewise (Griffin, 2020). When designing tools to counter hate, we must carefully consider the consequences of those tools being used by authoritarian states.

Furthermore, in moving from public-facing social media websites to Telegram and other less widely accessible sites, extremist networks have shown an ability to rapidly migrate to new platforms (Urman & Katz, 2020). If surveillance or deplatforming on Telegram inconveniences them, many hate groups will simply move to a new service such as Signal or Gab. As such, ameliorating hate cannot be left to any one service. Rather, tech companies should design their products to prevent hateful content multiplying inside defined online spaces like Facebook groups as well as hate movements from 'breaking out' into public social media. In particular, since harassment can be coordinated on one service and executed on another, all social media companies should design their products to be harassment-resistant. This is not simply a question of stricter moderation, but will likely require additional safety features and changes to existing aspects of these platforms. If the structures which previously enabled online hate remain the same, it will flare up again whenever it escapes containment.

Recommendations for Government and Civil Society

Governments at the federal, state and local levels need to adopt more comprehensive deradicalization initiatives. Since these groups are so difficult to

eradicate entirely, deradicalization initiatives should be focused on preventing vulnerable people from joining extreme groups online, deradicalizing users, regardless of whether they are members of online hate groups, and to prevent radicalization to the point of violence. Deradicalization is a well-studied subject in the context of other ideologies and conflicts, and those strategies can be applied to antisemitic and anti-Black groups. Civil society organizations can provide pathways to deradicalization, reducing the danger that hateful discourse will cross over to extremist violence.

Key stakeholders, from policy makers to community leaders and other influencers, need to use their bully pulpits firmly, persuasively and consistently to debunk hate-adjacent conspiracy theories, and attempt to do so in ways that are most effective, including to those vulnerable to radicalization. Engaging influential voices, in entertainment, sports and music recently have become more likely to join such campaigns. Conspiracy theories, such as those surrounding George Soros or Holocaust Denial, are possible gateways to radicalization - it is certainly possible for a user to go from merely partisan public pages to private extremist groups by exploring such theories.

Given the extraordinary risk posed to individuals, communities and democratic guardrails, Government must continue to play a role -- through public hearings and other means -- with respect to public oversight and the transparency and accountability of tech companies, particularly with regard to the development and enforcement of cyberhate policies. Such a role, however, must be consistent with constitutional prohibitions against Government censorship and overreach.

Finally, researchers, journalists and others with relevant expertise should map and monitor extremist groups and relevant government programs should formally include diverse civil society and vulnerable community participation and input. While it is ultimately up to the government and law enforcement agencies to monitor and prevent potential terrorists, mapping extremist groups allows us to understand how they introduce their messages into public discourse and how

they recruit new members. Rather than allowing hate to grow quietly, away from the public eye, we must ensure that successes are built upon.

BIBLIOGRAPHY

ADL. (2016). ADL Audit: Anti-Semitic Assaults Rise Dramatically Across the Country in 2015. Anti-Defamation League. https://www.adl.org/news/press-releases/adl-audit-anti-semitic-assaults-rise-dramatically-across-the-country-in-2015

ADL. (2018a). Quantifying Hate: A Year of Anti-Semitism on Twitter. Anti-Defamation League. https://www.adl.org/resources/reports/quantifying-hate-a-year-of-anti-semitism-on-twitter

ADL. (2018b). The Antisemitism Lurking Behind George Soros Conspiracy Theories. Anti-Defamation League. https://www.adl.org/blog/the-antisemitism-lurking-behind-george-soros-conspiracy-theories

ADL. (2019). White Supremacists Embrace "Accelerationism." Anti-Defamation League. https://www.adl.org/blog/white-supremacists-embrace-accelerationism

ADL. (2020a). After Long Fight, ADL is Relieved at Facebook Announcement That It Will Remove Holocaust Denial Content. Anti-Defamation League. https://www.adl.org/news/press-releases/after-long-fight-adl-is-relieved-at-facebook-announcement-that-it-will-remove

ADL. (2020b). Audit of Antisemitic Incidents 2019. Anti-Defamation League. https://www.adl.org/audit2019

Andrew, G. (2020, June 18). Russia lifts ban on private messaging app Telegram. The Independent. https://www.independent.co.uk/news/world/europe/telegram-russia-ban-lift-messaging-app-encryption-download-a9573181.html

Arif, A., Stewart, L. G., & Starbird, K. (2018). Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 20:1–20:27. https://doi.org/10.1145/3274289

Beirich, H., Buchanan, S., & SPLC. (2018). 2017: The Year in Hate and Extremism. Southern Poverty Law Center. https://www.splcenter.org/fighting-hate/intelligence-report/2018/2017-year-hate-and-extremism

Chandaluri, R. K., & Phadke, S. (2019). Cross-Platform Data Collection and Analysis for Online Hate Groups. https://vtechworks.lib.vt.edu/handle/10919/96292

Day, E. (2015, July 19). #BlackLivesMatter: The birth of a new civil rights movement. The Observer. https://www.theguardian.com/world/2015/jul/19/blacklivesmatter-birth-civil-rights-movement

Europol. (2019, November 25). Europol and Telegram take on terrorist propaganda online. Europol. https://www.europol.europa.eu/newsroom/news/europol-and-telegram-take-terrorist-propaganda-online

Gallagher, R. J., Reagan, A. J., Danforth, C. M., & Dodds, P. S. (2018). Divergent discourse between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. PLOS ONE, 13(4), e0195644. https://doi.org/10.1371/journal.pone.0195644

Koebler, J., & Haskins, C. (2019, May 16). Thousands of Facebook Groups Go Secret in Fear of the Great "Zuccing." https://www.vice.com/en/article/597dnb/thousands-of-facebook-groups-go-secret-in-fear-of-the-great-zuccing

Mezzofiore, G., & Polglase, K. (n.d.). White supremacists openly organize racist violence on Telegram, report finds. CNN. Retrieved October 20, 2020, from

https://www.cnn.com/2020/06/26/tech/white-supremacists-telegram-racism-intl/index.html

Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling Social Movements Through Social Media: The Case of Black Lives Matter. Social Media + Society, 4(4), 2056305118807911. https://doi.org/10.1177/2056305118807911

Parker, K., Horowitz, J. M., & Anderson, M. (2020, June 12). Majorities Across Racial, Ethnic Groups Express Support for the Black Lives Matter Movement. Pew Research Center's Social & Demographic Trends Project. https://www.pewsocialtrends.org/2020/06/12/amid-protests-majorities-across-racial-and-ethnic-groups-express-support-for-the-black-lives-matter-movement/

Potok, M. (2015, October 27). Carnage in Charleston. Southern Poverty Law Center. https://www.splcenter.org/fighting-hate/intelligence-report/2015/carnage-charleston

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. European Journal of Communication, 35(3), 213–229. https://doi.org/10.1177/0267323120922066

Thompson, A. C., ProPublica, Winston, A., & BondGraham, D. (2020, October 19). Racist, violent, unpunished: A white hate group's campaign of menace. ProPublica. https://www.propublica.org/article/white-hate-group-campaign-of-menace-rise-above-movement?token=aA9lRugYQlkPCYvk_Z6E9tDkP_5srI8X

Urman, A., & Katz, S. (2020). What they do in the shadows: Examining the far-right networks on Telegram. Information, Communication & Society, 0(0), 1–20. https://doi.org/10.1080/1369118X.2020.1803946

Wagner, K. (2020, September 23). Facebook Tool Faulted for Lapses in Finding Voter Misinformation. Bloomberg.Com.

https://www.bloomberg.com/news/articles/2020-09-23/facebook-tool-for-finding-voting-misinfo-falls-short-study-says

Woolley, S., & Joseff, K. (2018). Computational Propaganda, Jewish-Americans and the 2018 Midterms: The Amplification of Anti-Semitic Harassment Online. Anti-Defamation League. https://www.adl.org/resources/reports/computational-propaganda-jewish-americans-and-the-2018-midterms-the-amplification

Zannettou, S., Finkelstein, J., Bradlyn, B., & Blackburn, J. (2019). A Quantitative Approach to Understanding Online Antisemitism. ArXiv:1809.01644 [Cs]. http://arxiv.org/abs/1809.01644

ABOUT THE AUTHORS

ADL's Center for Technology and Society commissioned this study with Samuel C. Woolley, Director of the Propaganda Research Team at the Center for Media Engagement in the University of Texas at Austin and a former ADL Belfer Fellow, who in turn collaborated with Maggie Engler, a Senior Data Scientist at the Global Disinformation Index.

Samuel C. Woolley, PhD:

Dr. Samuel Woolley is a writer and researcher focused on how emerging media technologies are leveraged for both freedom and control. His new book, The Reality Game: How the Next Wave of Technology Will Break the Truth (PublicAffairs), explores how tools from artificial intelligence to virtual reality are being used in efforts to manipulate public opinion and discusses how society can respond. Dr. Woolley is an assistant professor in the School of Journalism and the School of Information (by courtesy) at the University of Texas (UT) at Austin. He is the program director of the propaganda research lab at UT's Center for Media Engagement and co-director of disinformation research for the UT "Good Systems" grand challenge—a university wide project exploring ethical AI design. His writings on tech, propaganda, and policy have been published by the National

Endowment for Democracy, the Anti-Defamation League, the Brookings Institution, the Stanford Hoover Institution, USAID, and the German Marshall Fund. He is a regular contributor to Wired, the MIT Technology Review, Slate, and a number of other publications. His research has been featured in the New York Times, the Wall Street Journal, and the Financial Times. He is the former director of research of the Computational Propaganda Project at the University of Oxford and the Founding Director of the Digital Intelligence Lab at the Institute for the Future in Palo Alto, CA. He is a former Belfer fellow at the Center for Technology and Society at the Anti-Defamation League and a former research fellow at the German Marshall Fund of the United States, Google Jigsaw, the Tech Policy Lab, and the Center for Media, Data and Society at Central European University. He has past academic affiliations with CITRIS at UC Berkeley and the Oxford Internet Institute at the University of Oxford. His PhD is from the University of Washington in Seattle. He tweets from @samuelwoolley.

Mark Kumleben, MA:

Mark Kumleben is a research affiliate with the Propaganda Team at the Center for Media Engagement at the University of Texas at Austin. He was formerly a research fellow with the Digital Intelligence Lab (DigIntel) at Institute For the Future in Palo Alto, CA. His research focuses on computational propaganda and disinformation in varied contexts, from election security to public health. Based in Chicago, Mark is an advocate for science communication and sound technology policy. His academic research includes the philosophy of artificial intelligence, and the nature and consequences of current developments in AI. Mark holds an M.A. in Politics and Economics from Claremont Graduate University, and a B.A. in Philosophy from the University of Chicago.

Maggie Engler, MS:

Maggie Engler is the lead data scientist for Global Disinformation Index (GDI), a nonprofit that aims to disrupt the business model of disinformation through

detection and demonetization. She is also an Assembly Fellow at the Berkman Klein Center for Internet & Society at Harvard University. Maggie's research is focused on applying statistics and machine learning to mitigate abuses in the online ecosystem, including disinformation, harassment, and fraud. Prior to GDI, Maggie spent several years in cybersecurity, most recently working on identifying anomalous login attempts at Duo Security. She has worked on a range of information security problems including malware classification and risk assessment in both the private and public sectors. Maggie holds a B.S. and an M.S. in Electrical Engineering from Stanford University, with a concentration in signal processing and a Notation in Science Communication with distinction.

DONOR RECOGNITION

This work is made possible in part by the generous support of:

The Robert Belfer Family

Dr. Georgette Bennett

Catena Foundation

Craig Newmark Philanthropies

The David Tepper Charitable Foundation Inc.

The Grove Foundation

Joyce and Irving Goldman Family Foundation

Horace W. Goldsmith Foundation

Walter & Elise Haas Fund

Luminate

One8 Foundation

John Pritzker Family Fund

Quadrivium Foundation

Righteous Persons Foundation

Riot Games

Alan B. Slifka Foundation

Amy and Robert Stavis

Zegar Family Foundation