Best Practices to Combat Antisemitism on Social Media

Research Report to the U.S. Department of State
Office of Religion and Global Affairs

July 2017

INSTITUTE FOR THE STUDY OF CONTEMPORARY ANTISEMITISM
INDIANA UNIVERSITY
Robert A. and Sandra S. Borns Jewish Studies Program
Bloomington

Acknowledgments

The report was prepared for the U.S. State Department as part of the Diplomacy Lab initiative, a cooperation between Indiana University and the U.S. Department of State and its Office of Religion and Global Affairs. It is based on research by students of the Spring 2017 course "Contemporary Antisemitism in Sociohistorical Perspectives" at Indiana University under the guidance of Günther Jikeli with the assistance of David Axelrod.

Contributing researchers are Harold Aaronson, David Axelrod, Ella Berry, Alison Borowsky, Günther Jikeli, Maia Katz, Jacob Levy, Danielle Maxwell, Luca Nemes, Robert Rajfer, Holden Rosenthal, Abraham Shapiro, Michaela Simon, Jenna Solomon, Jack Weinstock, and Bryan Woods.

We thank former U.S. Special Envoy to Monitor and Combat Anti-Semitism (SEAS) Ira Forman, Jerome E. Copulsky, Holly R. Huffnagle, and Helene M. Kessler at the Office of Religion and Global Affairs for their encouragement and support throughout this study and for their inspiring and helpful comments. We thank Indiana University scholars Olga Kalentzidou, Michael W. Hamburger, and Lee A. Feinstein, dean of the School of Global and International Studies, for bringing the Diplomacy Lab initiative to Indiana University and for their support and encouragement throughout the project.

We are grateful to Alvin H. Rosenfeld, director of the Institute for the Study of Contemporary Antisemitism (ISCA), for his help in enabling this research at Indiana University. We thank David Axelrod and Morgan Morales for their help in editing this report.

Contact:
Günther Jikeli, Justin M. Druck Family Scholar
Institute for the Study of Contemporary Antisemitism (ISCA)
Borns Jewish Studies Program
Indiana University, Bloomington
https://isca.indiana.edu/

Table of Contents

Executive Summary

With the advent of the Internet, antisemitic messages are disseminated more quickly and widely than ever before, and often go unchallenged. Veritable norms of antisemitism have been established in some social media circles. Within these circles, those who disagree with the antisemitic norm and venture into the conversation are ridiculed, attacked, or excluded, seriously impacting any ability to exert a positive influence on the conversation. Even more troubling, antisemitic messages often include incitement to violence and are contextualized in big-picture world-order ideologies that are bolstered by alternative news sources and "alternative facts."

This study looks at attempts to combat antisemitism on social media, and utilizes a survey with non-governmental organizations (NGOs) from Europe, Israel and North America which have been working in the field. Additionally, analysis of antisemitic posts and their disseminators and observations on interactions with disseminators of such posts provide us with additional background information to use in the development of new strategies to combat antisemitism online.

Attempts to work with social media and online platforms to **take down antisemitic content** have shown some success in recent years. However, major obstacles remain, and too many hate messages, including calls for violence, are never taken down. A clearer legal framework and closer cooperation between social media and online platforms, NGOs, and authorities will be necessary to take down antisemitic content in a timely manner without undue restrictions on freedom of speech. In view of European case law and legislative changes in some European states, such as Germany and France, IT companies will increasingly be held responsible for content that is disseminated on their platforms. NGOs can play an important role in flagging antisemitic content and in providing training to social media providers in correctly identifying antisemitic content, but ways have to be found so that the bulk of the work and financial burden does not remain with NGOs. As of now, it is still common to have social media employees dismiss user reports on blatantly antisemitic content. To help fix this situation, providers need to invest in training and technical solutions for monitoring hateful content. Terms of service that do restrict the dissemination of such messages need to be enforced. Policy makers can pave the way for effective regulations in countries where hate speech is illegal, and they can encourage the enforcement of terms of service that restrict hate messages. Social media interactions defy borders but effective measures in combating antisemitism need to take into account the regulatory framework, traditions, and forms of antisemitism, that are specific to each country and their constituent demographics. Social media users often utilize hyperlinks to antisemitic content on websites such as YouTube and blogs. A comprehensive approach that takes into account both social media and website content is therefore necessary.

While there seems to be a consensus among NGOs in the field that extreme antisemitic messages should not remain published on social media and taking such content down should be a priority, not all antisemitic content can be tackled in such a manner. **Counter-narratives will have to complement these efforts and reduce the negative impact of antisemitic messages that are not taken down**. This can be done by directly challenging antisemitic messages, and by calling out the disseminators for their hateful rhetoric. Another proactive method involves disseminating positive narratives or non-biased facts about Jewish people and Israel. However, counter-narratives face a number of challenges to being effective, such as reaching the target audience, being convincing, and not counter-productively giving antisemitic messages greater visibility than would have resulted if the messages where simply ignored. As current counter-narrative messages are done manually, they are time consuming and labor intensive, and, if done by individual users, expose them to attacks.

Our research about major disseminators of antisemitic messages in English shows **three main groups of disseminators** whose ideologies sometimes overlap: 1) white supremacists; 2) users who seem to be obsessed with Israel and who often consider themselves anti-Zionists and claim to be pro-Palestinian; and 3) users who might only use fragments of supremacist or anti-Zionist ideologies but who believe in a wide array of conspiracy theories. Anti-Zionist conspiracy theories are often a common denominator, although direct interaction between white supremacist and anti-Zionist disseminators of antisemitic messages appears limited.

Disseminators of the most influential antisemitic messages in terms of reach and re-posting tend to post such content regularly, peaking during relevant current events involving Jews or Israel. Closing accounts of these disseminators would be an effective means to reduce antisemitic content online even if they will be recreated under different names. Our observations of attempts to engage critically with disseminators of antisemitic posts show a number of challenges for counter-narrative efforts. The majority of disseminators simply ignore critical responses. Others double-down on their hateful messages and attack those who question or criticize their antisemitic posts. Antisemitic Twitter users react more aggressively and rudely than Facebook users, possibly due to a greater level of anonymity on Twitter. Very few disseminators of antisemitic posts feel the need to justify their position, and only exceptionally do they excuse themselves for using antisemitic tropes or insults.

Introduction

The spread of antisemitic messages on social media and the Internet, and its rise in recent years has worried many users, anti-racism organizations, and Jewish advocacy groups which have brought this to the attention of the wider public. Antisemitic content is more readily available than ever before, and has become part of the norm in many social media networks. This has been reinforced by social media circles wherein tailored newsfeeds and advertisements, and closed circles of "friends" make interaction with non-like-minded people scarce. Lawmakers and IT companies have increasingly shown that they are committed to dealing with the problem in a larger context by framing the issue of online antisemitism as part of the overarching problem of online hate speech and incitement to violence or terrorism.

Traditional media cultures and regulations need to be adapted and redeveloped for newer media platforms. Freedom of speech is and will remain a fundamental and inviolable principle for the media in democratic societies. It is however restricted by laws pertaining to copyrights, libel, and incitement to violence, or "illegal hate speech." These restrictions, however, differ from country to country and between private and public spheres.

What rules apply for messages posted on social media?

Hate speech can itself become a factor that stifles free speech. A survey from 2016 in all countries of the European Union shows that the majority of people surveyed who follow debates on social media have witnessed or experienced examples of abuse, hate speech, or threats. For almost half of those surveyed, these experiences make them hesitant to engage in online debates.[1]

IT companies that provide social media and online services have been pressured to do more to remove hateful messages from their platforms. Voluntary agreements, such as the Code of Conduct on illegal online hate speech, announced by the European Commission together with Facebook, Twitter, YouTube, and Microsoft in May 2016, have not brought the desired results, despite major improvements during the first six months of 2017.[2] The European Union and individual countries have now taken measures to make social media companies such as Facebook, Twitter and Google's YouTube more responsible for content

[1] European Commission. "Media Pluralism and Democracy," Special Eurobarometer 452, November 2016.

[2] See the press release of the European Commission, June 1, 2017, on "Code of Conduct on countering illegal online hate speech 2nd monitoring," http://europa.eu/rapid/press-release_MEMO-17-1472_de.htm (accessed June 2, 2017) and the fact sheet "Code of Conduct on countering illegal hate speech online: One year after," published by the European Commission in June 2017, http://ec.europa.eu/newsroom/document.cfm?doc_id=45032 (accessed June 12, 2017).

that is published on their platforms with threats of fines if illegal content is not removed after it has been signaled to them. However, the current praxis, that IT companies rely on users to flag illegal messages of hate that they can then review and remove, has been criticized because it puts the greater part of the burden on individual users or advocacy groups.

IT companies also increasingly face lawsuits that try to prove that platforms where incitement to hate crimes and terrorism was published bear some responsibility for the subsequent crimes. Last but not least, companies have become increasingly reluctant to pay for advertisements on social media platforms if they fear that their mainstream brand is displayed alongside hateful posts.

However, even if the rising pressure, closer cooperation between IT companies, lawmakers and NGOs in the field, and improved algorithms lead to better removal rates of hate messages, this can only reduce the visibility and impact of the worst online messages. Counter-narratives will have to expose and counter false accusations, conspiracy theories, prejudices and negative sentiments against Jews and other groups.

What are effective counter-narrative strategies? How can major challenges be overcome, such as reaching the target audiences and penetrating their social bubbles, being convincing, and not granting antisemitic messages more third-party attention than would have resulted if the messages where never challenged? And how can counter-narratives at least in part be automated to cope with the vast numbers of antisemitic messages that need to be countered?

Definition of Antisemitism

For the purpose of this study we use the definition of the U.S. State Department that refers to the Working Definition of Anti-Semitism by the European Monitoring Center on Racism and Xenophobia. The International Holocaust Remembrance Alliance with 31 member states, including the U.S. and 24 EU member countries, uses a similar definition.[3]

We use the non-hyphenated spelling of “antisemitism” that has become more and more common in academic research and might help prevent the misunderstanding that antisemitism is simply opposition to “Semitism.” We use the term “anti-Zionist” as a self-declaratory term because many users apply this term to themselves. We do not consider all forms of anti-Zionism as antisemitic but we apply the definition of antisemitism relative to Israel (see below) to anti-Zionism. Anti-Zionism is a strong indication for antisemitic attitudes if it denies

[3] https://holocaustremembrance.com/media-room/stories/working-definition-antisemitism (accessed June 29, 2017).

the Jewish people their right to self-determination and if it denies Israel the right to exist.

The U.S. State Department's Working Definition states: [4]

"Anti-Semitism is a certain perception of Jews, which may be expressed as hatred toward Jews. Rhetorical and physical manifestations of anti-Semitism are directed toward Jewish or non-Jewish individuals and/or their property, toward Jewish community institutions and religious facilities."
WORKING DEFINITION of ANTI-SEMITISM
by the European Monitoring Center on Racism and Xenophobia

CONTEMPORARY EXAMPLES of ANTI-SEMITISM

- Calling for, aiding, or justifying the killing or harming of Jews (often in the name of a radical ideology or an extremist view of religion).
- Making mendacious, dehumanizing, demonizing, or stereotypical allegations about Jews as such or the power of Jews as a collective – especially but not exclusively, the myth about a world Jewish conspiracy or of Jews controlling the media, economy, government or other societal institutions.
- Accusing Jews as a people of being responsible for real or imagined wrongdoing committed by a single Jewish person or group, the state of Israel, or even for acts committed by non-Jews.
- Accusing the Jews as a people, or Israel as a state, of inventing or exaggerating the Holocaust.
- Accusing Jewish citizens of being more loyal to Israel, or to the alleged priorities of Jews worldwide, than to the interest of their own nations.

What is Anti-Semitism Relative to Israel?

EXAMPLES of the ways in which anti-Semitism manifests itself with regard to the state of Israel, taking into account the overall context could include:

DEMONIZE ISRAEL:

- Using the symbols and images associated with classic anti-Semitism to characterize Israel or Israelis.

[4] https://www.state.gov/s/rga/resources/267538.htm (accessed June 29, 2017).

- Drawing comparisons of contemporary Israeli policy to that of the Nazis.
- Blaming Israel for all inter-religious or political tensions.

DOUBLE STANDARD FOR ISRAEL:

- Applying double standards by requiring of it a behavior not expected or demanded of any other democratic nation.
- Multilateral organizations focusing on Israel only for peace or human rights investigations.

DELEGITIMIZE ISRAEL:

- Denying the Jewish people their right to self-determination, and denying Israel the right to exist.

However, criticism of Israel similar to that leveled against any other country cannot be regarded as anti-Semitic.

Methods

In the first component of this study, we sent out a survey to non-governmental organizations (NGOs) in the field, asking for their strategies for combating online antisemitism and their views regarding counter-narrative efforts. You will find the survey questionnaire in the annex. Representatives from 17 organization from Australia, Belgium, Czech Republic, Israel, France, Germany, Netherlands, Sweden, United Kingdom, United States, and Europe kindly took the time to respond.[5] They provided us with detailed information on their work on combating online antisemitism, and pointed out challenges and strategies currently used in combating online antisemitism.

Secondly, we searched social media platforms for antisemitic posts, concentrating on Twitter, and analyzed the background of the most influential disseminators.[6] We used the social media analysis tools of Keyhole, Netlytic, and Digital Methods Initiative - Twitter Capture and Analysis Toolset (TCAT) to find antisemitic posts falling within six typological categories based off the State Departments definition of antisemitism: conspiracy theories, Holocaust denial, antisemitic forms of anti-Zionism/anti-Israelism, insults, physical tropes, and

[5] Some of the respondents and organizations wished to stay anonymous and we therefore do not provide a list of organizations that responded. However, we contacted the organizations that can be found in various reports in the references in addition to a few others. The response rate was over 70 percent.

[6] The most influential disseminators were defined as by the numbers of likes and shares they earned for the antisemitic posts that we were looking at.

religious tropes. The searches were done between April 7 and May 10, 2017, with search terms specialized to each category. [7] We utilized Keyhole's ability to track the most influential users in terms of their posts' engagement, defined by the number of likes and shares received by the posts containing our keywords.[8] We proceeded to select the top five most influential users for each keyword. In one category, Holocaust denial, we also analyzed the top 30 profiles of users who sent out the largest number of messages containing the keyword. After profiling the top five individuals from the 'engagement' category for each search-term, we categorized them according to ideological, political, and demographic guidelines to examine if there are any consistencies, patterns, or overlap among the disseminators.

Thirdly, we responded to antisemitic messages on Twitter and Facebook, engaged in debates with some of the disseminators, and observed the reactions. Our response approaches to antisemitic messages fell within four categories of counter-narrative responses: declarative statements such as: "This is antisemitic!", emotional responses from a victim's perspective, humorous memes to ridicule antisemitic statements, and data-based arguments. On Twitter, we experimented with a bot, trying to find ways by which disseminators of antisemitic messages would receive an automated or semi-automated critical response. We analyzed both our interactions with disseminators of antisemitic posts, as well as their profiles.

---

[7] We tested a number of different keywords, aiming for terms that yield a large amount of messages in a certain time period and a large percentage of those being antisemitic. We chose the following key-terms: "Jew Media" OR "Jewish Media", "Jewish money" OR "Jewish Bankers" OR "Jew Banks", "Jewish Power" OR "Jewish control", "evil Talmud", "Holohoax", "Zionist Nazi", "Zionist lies" OR "Zionist propaganda", "Jew nose", "Kikes", and "Dirty Jew." Searches with these keywords generated a large percentage of antisemitic posts. However, they also captured some posts that expressed opposition to such content.

[8] A comparison of the data obtained by Keyhole to data obtained by Netlytic confirms that data by both tools are fairly similar to one another as shown in the case of the keyword "Holohoax" during the date range of April 30th to May 9th. The sites agree on 4 of the top 5 propagators, and 6 of the top 10. While this may not appear to be an excellent match, when one accounts for the small standard deviation in post numbers after the top six users, along with the fact that Keyhole counted nearly 270 more tweets than Netlytic (1097 to 830 respectively), the divergence after the top five becomes expected (a difference of just one tweet can change the order ranking). Though the difference in total "Holohoax" tweets counted by each site is problematic, this might be the result of the sites having different approaches to counting tweets of users that had been removed by twitter administrators, as the keyword "Holohoax" was propagated by a relatively high number of users that had their accounts suspended for violating Twitter community standards. While we cannot be sure of either site's precision at this time without knowing more about the algorithms they employ, we do see enough similarities to identify trends worth exploring in future studies.

Types of Antisemitism and Antisemites on Social Media

Most of the NGOs that responded to our survey claimed "traditional" or "classic" antisemitism as the most prevalent form found on social media platforms. Stereotypes include the idea that Jews control the financial world, media and Hollywood, and are engaged in an attempt to destroy traditional or nationality centered societies. One popular theme within classic forms of antisemitism is the trope that Jews lobby to open the borders (of Europe and America) in order to "exterminate" the "white race." This is often in line with Nazi propaganda that seeks to demonize and dehumanize Jews by portraying Jews as systemically perverting the media to destroy the morality of the "white race," or by portraying Jews as pedophiles and sexual deviants. White supremacist antisemitism has had a resurgence. In central and eastern European countries, there are, in addition to users from the far right and neo-Nazis, a rising number of neo-Stalinists who disseminate antisemitic propaganda. These groups share anti-liberal, anti-Western and pro-Russian attitudes, and also peddle many conspiracy theories. Many of the organizations we surveyed also noted a rise in what can be termed as the "new antisemitism" directed against Israel, which attempts to portray Israelis or Zionists as the "new Nazis," often in conjunction with news reports of and developments in the Israeli-Palestinian conflict. Many of the purveyors of the "new antisemitism" claim to be "pro-Palestinian." These forms of antisemitism can be expressed in outright calls for violence against Jews, or they can be subtle but nevertheless lead to exclusion and vilification of Jews and of those who do not agree with a norm of demonization of Israel.

Another troubling trend for many NGOs is the rapid growth of Holocaust denial. In addition to hardcore Holocaust denial, somewhat softer forms include the accusation that the Holocaust has been exaggerated by the Jews in order to create Israel, or that Holocaust remembrance has been enforced in order to distract from Israel's alleged monstrous atrocities. Incitement to violence against Jews and Israelis, including incitement to terror has also been reported as a rising problem. According to our survey responses, the most frequently mentioned disseminators of antisemitic posts in English are neo-Nazi outlets such as The Daily Stormer, Stormfront, Iron March, David Duke[9], forums with many white supremacists and neo-Nazi participants such as 4chan/pol, and also users who claim to speak in the name of Christianity, such as Brother Nathanael. Posts by virulent anti-Zionist outlets such as Electronic Intifada are also widely disseminated, particularly in times of crisis in the conflict between Israel and Palestinians.

Our analysis of Twitter messages in English from April 7 to May 10, 2017 showed that the most influential disseminators of antisemitic messages in terms of numbers of likes and retweets were sent by white nationalist, often self-

[9] David Duke has now begun to refer to himself as a civil-rights advocate for white people

identified or clearly affiliated 'alt-right' individuals. Based on their profiles and the ideology they promote via their accounts; they can be situated within recent waves of anti-globalist support gaining traction across the world. These individuals from the far right were extraordinarily and proudly antisemitic, seeking to promote their ideology and change the minds of other less "enlightened" individuals. Antisemitic accusations and Holocaust denial is presented as the truth and an expression of free speech pushing back against the Jewish controlled "mainstream media." As it relates to Israel, these users showed a number of anti-Zionist conspiracy theories, fear of Zionist manipulation of the world, and particularly (for North American users) United States security interests. They were often supporters of Donald Trump, and would also voice support for right-wing or autocratic leaders such as Marine Le Pen, Vladimir Putin, Bashar al-Assad, and Victor Orban. The majority of the white supremacist users, fearing a "white genocide," also spread hatred against immigrants, Muslims, the LGBTQ community, and people with disabilities. Some even believe that these minorities, including Muslims, are being instrumentalized by the Jews to threaten the "white race."

Another group of disseminators of antisemitic messages self-identifies as "pro-Palestinian" or "anti-Zionist," but may be more accurately referred to as anti-Israel, utilizing stereotypical images of Palestinians as victims only while casting Israel as an evil entity. These disseminators can be situated largely within the political far-left. They frequently post in ways that demonize Israel by extending traditional antisemitic tropes to Israel. Extreme cases are apologetic toward terror against Israeli or Jewish targets or even endorse terrorism. Interestingly, these individuals appear to peddle some similar conspiracy theories to their white nationalist counterparts, that is conspiracy theories about Israel running the world and being at the root of wars and crisis related to ISIS and the Syrian refugees. The most common point of divergence between these two groups, however, is that this group of far left antisemites did not propagate hate of minorities besides Jews. Instead, its focus commonly remained on Israel and Jewish-centered conspiracy theories. There was also less Holocaust denial from this group, though not too uncommon. When it did occur, it more frequently took the form of minimization or seeing Holocaust remembrance efforts as an attempt to divert attention from more 'relevant' issues, such as the plight of the Palestinians. Thus, for these users, Holocaust remembrance was seen as part of a zero-sum competition for the world's attention. Some "pro-Palestinian" users only occasionally sent clearly antisemitic messages. These users were generally far more concerned with Palestinian human rights and Israeli political issues, but that commonly provided a simple segue to fear of Jewish and Zionist control of the media, wealth, and power. Some frequently reposted messages/pictures from well-known sites and figures such as electronicintifada.net or Carlos Latuff.

A third group of influential disseminators of antisemitic posts is neither obsessed with Israel nor can they be identified as extreme right, but they nonetheless believe in antisemitic conspiracy theories, mistrust the media, and sus-

pect that Jews and Israel run the media. Many of these antisemitic users lack a coherent or consistent political ideological framework, even though they may identify as a part of a broader movement. This makes it exceedingly difficult to categorize them. Often, the only true constant is their antisemitism. Many of these users only occasionally utilize anti-Jewish stereotypes or insults.

Most disseminators of antisemitic messages in all the categories examined appear to be male, but this pattern is least pronounced among the "pro-Palestinian" category.

Case Study on Online Holocaust Deniers: "Holohoax"

The popular term "Holohoax" stands for outright Holocaust denial of a particularly mocking nature and is used mostly, but not exclusively, by users whose world views can be described as far right, alt-right, or Neo-Nazi. We examined the top 30 propagators of tweets containing the keyword "Holohoax" as calculated by Keyhole for the time period of April 10th to May 10th. In this analysis, the top thirty users were ranked only by number of posts containing the keyword, and not by other markers of influence and reach. For each of these accounts, we identified their apparent political affiliation (far left/ far right/ moderate/ unclear), posting patterns, national origin, number of posts containing "Holohoax" for the examined period, and number of followers. For accounts that were suspended (7 out of 30 at the time the data was gathered), the number of posts prior to suspension, as well as quantity of followers was noted. Other data was unavailable for these accounts. Of the 23 active accounts, only one account, "ISRAEL BOMBS BABIES" @Col_Connaughton was almost exclusively concerned with the Israeli-Palestinian conflict, though this account was the second most prominent account in this period. Considering that most accounts are based in North America (assumption based on alt-right affiliation which is mostly confined to North America), it should be noted that this one is based in the UK. Four accounts specified being registered in the United States. Seven accounts were registered outside the United States: five in the United Kingdom, one in Greece, and one in Estonia. 19 accounts did not provide any data on the country of residence. Five of the 23 active accounts could not be easily placed into an ideological category, and they posted on a variety of topics—mostly conspiratorial content. 17 of the 23 active accounts can be identified as White racial nationalists, and among those, 7 (including the consistently and by a wide margin top poster) can be identified as belonging to the amorphous alt-right (identified by prevalence of markers including: memes, usage of Pepe the frog, anime, and citation of alt-right sites such as The Daily Stormer). Although all accounts contained anti-Zionist content, the top two posters were more frequent in their posts about Israel than the others which had greater topic variety. All accounts peddled conspiracy theories, many directly related to Jews. One of the accounts had Christian oriented material.

Most appear to be male users as only one user portrayed herself to be female. The three most active posters of "Holohoax" tweets also had the largest groups of followers (between 4,884 and 18,265). The user "Philosophile" @RealFKNNews was by far the most frequent poster of messages containing the term "Holohoax (398 posts from April 10th to May 10th), mostly sending out a meme produced some years ago by the now defunct website theforbiddentruth.net[10] "Holocaust Extortion for Dummies" with the caption "Holohoax for Dummies." This user regularly retweets a number of identical memes and messages within hours. The self-description of @RealFKNNews, "*Reporting on the REAL News, not MSM!,*" reflects the conspiratorial mindset and mistrust of media common among many such users. Users who posted about Holocaust denial often retweeted messages by other users such as @RealFKNNews, or references to websites such as The Daily Stormer. Some used references to Jewish and anti-Zionist Holocaust deniers such as Moshe A. Friedman, former Bishop Richard Williamson, or David Irving to support their claim that the Holocaust is a myth.

Efforts to Take Antisemitic Content Offline

NGOs have been in the forefront of flagging antisemitic content to Internet service providers and social media platforms so that they could remove such content, suspend the accounts, or take down the websites. Social media companies have developed different systems that allow users to flag hateful content, and they have established community standards delineating acceptable content, increasingly making it clear to their users that antisemitic and other hate-messages violate their terms of service. However, too many antisemitic messages, including calls for violence against Jews, remain online despite these measures.

The vast quantity of antisemitic messages and accounts is another clear obstacle. Accounts that are shut down are often recreated within minutes under different aliases. NGOs also reported of the danger of hostile sites migration to unregulated platforms. Other obstacles include the limited monitoring resources of NGOs and individual users, and the lack of human resources and expertise at social media platforms allocated to deciding what content should be taken down. This seems to be particularly true for smaller countries and less common languages. Social media providers do not seem to have sufficient human resource personnel in their complaints departments who understand the less prominent national and regional cultural contexts and languages—which would be necessary for understanding antisemitic messages that are specific to that context. Further, for ideological and financial reasons, social media plat-

[10] The website is still accessible but the content has been altered completely, with most content being unrelated advertisements texts, but featuring prominently articles on antisemitism.

form providers have been reluctant to block content and users, as they do not want to limit or censor their customers. Lawmakers in European countries increasingly pressure Internet service providers and social media platforms to remove hateful content. In the United States, due to its constitution and history regarding free speech, other measures will be more effective. [11] Providers can be encouraged to take reports about cyberhate seriously, enforce terms of service that do not allow for the dissemination of hate messages, provide more transparency of their efforts to combat cyberhate, and to offer user-friendly mechanisms and procedures for reporting hateful content.

However, the initial position taken by many social media platforms, that all responsibility for what is published on their platforms is solely the responsibility of the individual posters, has become increasingly indefensible; legally (at least in some countries), in terms of PR, and possibly most importantly, in the eyes of many social media users. Hateful messages can tarnish the brand name of these companies, which may lead to a decrease in income from advertisements. More recently, platform providers have been made legally responsible in some European countries for the content that is published therein, especially if they fail to take action after illegal content has been reported to them.

NGOs are actively helping in the process to take antisemitic content offline, or to reduce its visibility and impact by:

- Monitoring and reporting antisemitic posts with the help of volunteers, and developing tools for these purposes (such as fightagainsthate.com);
- Providing and maintaining an effective flagging process. Crowd reporting, such as done by theycant.com, seems to be particularly effective in sending a strong signal to social media providers;
- Monitoring of flagged antisemitic content to see if it actually gets removed;
- Helping social media providers acquire expertise in identifying antisemitic content;
- Encouraging an industry standard in defining online antisemitism;
- Encouraging norms of interaction between social media users which includes refraining from hate messages such as antisemitism;
- Establishing direct contacts to social media providers to point out particularly vicious messages and accounts to them that should be taken down;

[11] Under the First Amendment of the United States Constitution, "Congress shall make no law…abridging the freedom of speech, or of the press." Only narrowly defined categories of speech are unprotected by the First Amendment, such as responsive violence, defamation, obscenity, commercial speech, and true threats. Additionally, section 230 of the Communications Decency Act (1996) provides that: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."

- Raising public awareness of the dangers of online antisemitism;
- Publicly pressuring social media platform providers to become more proactive to remove antisemitic and other hateful messages;
- Making law enforcement and judicial authorities more aware of the problem if content might violate any law (such as libel or incitement to violence) and helping to build up expertise within police and judicial authorities;
- Lobbying policy makers to put more pressure on social media companies (most US organizations refrain from doing so because of the First Amendment)[12];
- Fighting antisemitism offline and in traditional media as this often feeds into social media.

Such efforts are mostly limited to a few of the largest social media providers, namely Facebook and Twitter, to YouTube as a source of countless antisemitic videos, and to antisemitic websites that also serve as resources and references for antisemites. Many cite a lack of funding and resources as the main obstacle to improving their work on social media. Smaller NGOs specialize in some of these actions by concentrating on one social media platform, or by working on campaigns for a certain time period or specific themes such as confronting Holocaust denial or antisemitic boycott campaigns against Israel. Most NGOs develop tools and strategies specific to the national context(s) they operate within which allows them to adapt their work to the local culture, language, and social and political offline networks. International cooperation does exist, but while some NGOs operate internationally, many NGOs stress that international and nation-wide cooperation along with direct cooperation with social media providers needs to be improved. Specifically, monitoring and reporting tools could benefit from increased cooperation among NGOs. Robin Sclafani (CEJI – A Jewish Contribution to an Inclusive Europe) sees "*a need for greater cooperation between law enforcement, public authorities, civil society and IT companies to address the problem of online hate in a more holistic way that is both reactive to emerging cases and issues, and proactive to achieve the cultural change needed to reduce the expression of hatred*."

Counter-Narratives

Counter-narratives used to combat online antisemitism are often viewed critically by NGO representatives. Some see it as an ineffective approach that social media platform providers are pushing in order to avoid taking action and responsibility. Others believe that counter-narrative strategies should complement attempts to take antisemitic content offline.

[12] There seems to be a shift also in the United States from seeing social media platform providers primarily as tech companies to seeing them increasingly as publishers. The latter implies some responsibility for the published content.

Michael Whine (Community Security Trust) points out that "*counter speech never has as much force or resonance as the original*" and Eliyahou Roth (They Can't) fears that "*even the counter speech could be destructive*" in face of certain vicious antisemitic accusations, and believes that the priority should be in taking such content offline. Jonathan Vick (Anti-Defamation League) says: "*In many ways the jury is still out on the effectiveness of counter-narratives. However, at the very least, anti-Semitism or cyberhate left without any resistance is perceived as tantamount to tacit endorsement*."

Only a few organizations are engaged in direct and systematic counter-narrative efforts online. NGOs have often reacted offline by publicly condemning examples of antisemitism on social media and seeking collaboration with politicians, public figures, and leading media outlets in the condemnation of antisemitism. In some cases, they reported that this had a positive effect on reducing the spread of certain antisemitic messages.

Counter-narratives can address antisemitic messages directly by questioning and rejecting them, or by taking to task the disseminators for their hateful messages. Counter-narratives can address antisemitic messages indirectly by providing positive narratives or cut and dry facts about Jewish people and Israel that undermine antisemitic tropes and conspiracy theories. This has the advantage of not validating antisemites by directly engaging with them, and it can provide important resources for the future on current and past events around which antisemitic rumors are based.

However, counter-narratives face a number of challenges to be effective including the difficulty of reaching the target audiences and penetrating their social bubbles, being convincing, and not granting antisemitic messages more third-party attention than would have resulted if the messages where never challenged. Exposure of and direct responses to antisemitic messages also leads to greater visibility in search engine results. Lastly, most direct counter-narrative approaches require significant human resources to be effective.

Best practices that confront antisemitic messages directly include:

- The creation of counter-narratives tailored to confront the specific falsehoods and absurdities of their claims, while also providing training on how to monitor and counter antisemitic messages such as done by "Facing Facts!"[13] or "Get the Trolls Out!"[14] These approaches do not directly respond to specific antisemitic content on social media, but rather train other users to do so and provide argumentative resources.
- Sending rebuttals, memes or factual articles individually in response to posts. This has been done by countless individual users. However, there are three important limitations to this approach: 1) The number of users

[13] http://www.facingfacts.eu/ (accessed May 31, 2017).
[14] http://www.getthetrollsout.org/ (accessed May 31, 2017).

who actively oppose antisemitic messages is relatively small compared to the number of users who disseminate them. 2) The approach is usually limited to the social media circles of the individual users who oppose antisemitism. These circles are less likely to capture the greater bulk of antisemitic messages. 3) Those who oppose antisemitic messages in social media circles can be silenced or blocked by other user if they oppose (antisemitic) group norms.

- Semi-automated responses to antisemitic messages that signal to users and their followers that these messages fall outside social norms. There have been only a few but promising experiments with bots working in tandem with resource websites such as the project "Nichts gegen Juden" [Nothing against Jews] by the Amadeu Antonio Foundation in Germany.[15]

To our knowledge, more advanced methods that include the profiling and targeting of individual social media users, and sending them tailored messages and information, have not been attempted yet.

Our interactions with disseminators of antisemitic messages during this study showed that counter-narratives do not work when employed against ingrained antisemites, but they can in some instances be instrumental in helping relatively unbiased individuals recognize and reject online antisemitism.

Case Studies in Countering Antisemitic Messages

In four groups, we experimented with counter-narratives in direct response to disseminators of antisemitic messages on Twitter and Facebook. While three of the four groups focused on Twitter, each took a different approach to their research. Nevertheless, we had similar findings across the groups in regards to the types of users who disseminate antisemitic messages, which users respond to counter narratives, and the potential impact of counter-narratives.

In order to find antisemitic posts, we used a large variety of search terms and also looked manually through posts that had the keyword "Jew(s)" and its synonyms. It is relatively easy to find hardcore antisemitic messages from the extreme right, but more difficult to find messages containing subtler forms of antisemitism, and where antisemitic sentiments are concealed in insinuations. Also, a major challenge for all groups was to get disseminators of antisemitic messages to react to our counter-narrative responses. The majority simply ignored our responses.

The group which focused their research on Facebook faced a challenge in that much of

[15] http://www.nichts-gegen-juden.de/ (accessed May 31, 2017).

Facebook can be private and thus not visible to those who are not friends. They began by setting up two accounts with pseudonyms, Will Brown and Roger Smith. In creating these personas, they attempted to appear as politically neutral as possible in an effort to get people to engage with them. Roger Smith liked posts that were both pro-Palestinian and pro-Israel. They did not intend to create an extensive backstory for Roger Smith, but still aspired to appear as a legitimate user. Through both Facebook accounts, the students found that respondents tended to be left-leaning supporters of Palestine, or propagators of conspiracy theories. They found that while people would respond when they tried to engage them in conversation, these responses were often filled with conspiracy theories and relied heavily on antisemitic tropes. Respondents were also quick to justify their antisemitism, but their justification often relied on mistaken and false information. Some reactions by antisemitic users to our responses might have been provocations, including outright Holocaust denial which they used as a potential trigger to get an emotional reaction from us. We found that an effective way of starting a conversation with antisemitic users was to show an interest in their views.

The three groups working on Twitter also set up accounts with pseudonyms. One group utilized bots in order to respond to antisemitic posts; when an antisemitic tweet was posted, the bot would respond with a pre-written answer. That group searched Twitter for keywords such as "kikes" and phrases such as "fuck Jews," the former of which yielded the most results. Their strategy was to respond either with memes, a declarative statement, or articles and data-heavy responses. The memes garnered the fewest reactions. Yet this in itself does not belie the memes' potential effectiveness or ineffectiveness since memes largely ridicule the conspiratorial nature of the antisemitic posts with the aim of influencing third parties and not the original poster. While the bots could send out the initial response, the interactions needed to be followed-up manually in the case of a reaction. The human component was necessary to produce actual results and not be dismissed as a bot. Many of the users who reacted had comparatively newer Twitter accounts, having joined in 2016 or 2017. The reactions were aggressive and seemed designed to evoke an emotional response from us. Responses came almost immediately, and users were unwilling to change their opinions.

The second group to use Twitter initially tried using a set of pre-written responses, but found that those did not get much attention. Like the other groups, they found that they achieved better results when they reacted personally to the antisemitic users rather than relying on stock answers. It was necessary to seem interested in what the antisemites had to say rather than to directly attack what they had to say. They found responses from both left-wing and right-wing users. Again, tropes were commonly used as justification for antisemitic beliefs. Some users admitted openly that they were racist, perhaps utilizing the anonymity buffer provided by social media to avoid the social consequences of professing racist beliefs.

A third group tried to find common denominators between antisemitic Twitter users. They found that while some women replied, most users were white males. Most of the antisemitic users also held other prejudices, and many were prone to conspiracy theories. Like the Facebook users, they relied on antisemitic tropes aided by online resources such as YouTube videos. Many of the users seemed uninterested in engaging in conversation, but seemed to be interested in provoking an emotional response from us. Many seem to avoid interacting with people that do not signal that they share the same prejudices.

Among the posters, there were three types of responders: antisemites with justification, antisemites without justification, and antisemites that would not categorize themselves as antisemites or as hating Jews. The majority of the users contacted were in the first category of responders. Oftentimes they identified themselves as Nazis or neo-Nazis and justified their claims with Jewish stereotypes or generalizations they extrapolated from brief encounters with Jewish people. The users truly believed they were righteous individuals protecting the public against perfidious Jews who wished for the demise of America and the "white race" as a whole. Additionally, many of them seemed to hold other prejudices against women, LGBTQ community members, and immigrants. Some of the accounts with which we interacted were suspended during the research period. The second category of users provided little to no explanation for their beliefs, instead deciding to continue writing antisemitic messages as responses. They didn't believe that they needed a justification for their hatred of Jews– all Jews are evil and that is that. Lastly, the users who would not think of themselves as antisemitic despite peddling dehumanizing tropes and stereotypes about Jews, were usually strongly anti-Israel and anti-Zionist and believed in conspiracy theories pertaining to Israel.

All our research groups found it impossible to change the views of the hardline antisemites with whom they interacted. Those individuals were set in their beliefs and utilized social media to find likeminded people and information that reinforced their engrained beliefs. The individuals willing to engage often had "facts" and figures available, as well as memes and links to websites and YouTube videos to rebuff our points. Most of them responded immediately. Conspiracy theorists would say that they had alternative facts or histories which were the actually and absolute truth. Most would say they do not trust or believe "fake" mainstream narratives, the "liberal media," and narratives that had been forced on them their whole lives. However, these users referred to unreliable sources and posted fake quotes to demonize the State of Israel and Jews in general. This extended to out of context or outright invented passages supposedly from the Talmud.

It could seem that responding to antisemitic messages would be futile beyond research purposes. However, while trying to change these users' views and sentiments may be a fool's errand, it does present an opportunity to show third parties that such views are not shared by everyone and/or that antisemitic accusations are false and are expressions of hateful sentiments rather than one's

interpretation of facts. While the number of interactions was relatively low, many third parties saw our responses to the antisemitic messages.[16]

However, engagement with people who utilized antisemitic tropes or insults without harboring deep antisemitic sentiments was more fruitful. For instance, the group using bots on Twitter found that one user who used the word "kike" as part of her nickname, responded that she did not know that the term was offensive. She was willing to respond, was apologetic, changed her nickname, and even made that point clear to her followers.

Targeting such users who do not harbor hardened antisemitic beliefs would require using more specific, perhaps more complex, search terms or profiling. However, explaining to people why what they have said is antisemitic might enlighten them to not only stop using the antisemitic term, phrase, or joke, but also increase their awareness of the overt forms of antisemitism on social media. In essence, this may provide another strategy to undermining the narratives of the hardcore antisemites. It might not be possible to completely remove the extreme antisemitic narratives from social media, but it is possible to inhibit their ability to sway third parties.

However, the question of how best to combat antisemitism on social media requires additional research with longer and larger case studies.

Challenges for NGOs

Making more people aware of the presence and destructive potential of online antisemitism is one of the major challenges facing the organizations we surveyed. Most NGOs cited a lack of funding and personnel as the most limiting challenge. A majority reported that beyond such difficulties, the sheer volume of antisemitic content on the Internet is daunting, and convincing social media platforms and Internet service providers to more aggressively remove content of this nature is an ongoing battle marked by sporadic and uneven progress. In too many instances, the content of terminated accounts simply reappears later under new aliases with different usernames, which might be an indication that providers do not utilize all the available technical solutions to this problem. However, a number of NGOs reported successes in efforts to have providers take down tens of thousands of antisemitic messages, uncovering previously unknown antisemitic individuals and organizations, and creating situations in which public condemnation of online material becomes more powerful than the effects of the initial antisemitic post.

---

[16] One of our Twitter accounts had 156 interactions. The account saw 7,500 impressions within a month, meaning that 7,500 active Twitter users saw this account's posts.

When asking what the NGOs would suggest as the most effective approach given sufficient manpower and funding, most NGOs responded that massive monitoring in more languages, coupled with campaigns of awareness raising among the wider public are important first steps. They also said that it is crucial to continue calling out social media and Internet service providers for their relative passivity to date. Many NGOs reported a need for greater cooperation between law enforcement, public authorities, civil society and IT companies to address the problem of online hate in a more holistic manner that involves not only reacting to emerging cases and issues, but also includes proactive measures to achieve the cultural transformation needed to result in fewer expressions of hatred online. Lastly, current aspects of legal systems in a number of countries undermine the ability to stem the torrential flow of online hatred.

Conclusions and Recommendations

NGOs have played an important role in denouncing antisemitism online and in bringing this to the attention of lawmakers and the wider public, relating it to the overarching issues of hate speech and incitement to violence and terror. NGOs have also done important work in flagging antisemitic content to social media and Internet providers and in helping other users to do so. These contents usually violate the terms of service, regardless if they are illegal in some countries. However, NGOs do not have the resources to monitor and flag antisemitic messages comprehensively. This is particularly true for less prominent languages. Providers should offer user-friendly mechanisms and procedures for reporting hateful content and enforce their terms of service regarding the dissemination of hate speech.

It might be possible to increase the flagging rate of individual users, but due to the structure of social media, those who are sensitive to antisemitic content are less likely to see it than those who might be drawn into such messages. Thus, social media and Internet providers cannot rely on users and NGOs to monitor antisemitism on their platforms. They need to take active monitoring measures themselves if antisemitic messages or illegal hate messages are to be removed in a comprehensive manner. They also need to improve the supportive structure for flagging such content, take swift action when antisemitic messages are signaled to them, and provide more transparency on their efforts to combat cyberhate on their platforms.

Our research shows that many influential disseminators of antisemitic posts in terms of reach, likes, and reposts, send out antisemitic messages regularly, often using the exact same messages, whereas large numbers of antisemitic posts

garner low interest and a low level of virality online.[17] Closing influential antisemitic accounts and/or taking such posts and all correlating re-posts offline can significantly reduce the amount of antisemitic content. Intensified international cooperation between different NGOs, lawmakers, and IT companies in monitoring and flagging antisemitism online can help to improve efforts in monitoring and taking antisemitic content offline and prevent republishing under different names and migration of hateful content to unregulated platforms. This is particularly crucial in less prominent languages. Cooperation might also help smaller NGOs that specialize in monitoring online antisemitism to get the financial support that is needed to be effective.

Not all antisemitic content can be removed and the potential of counter-narratives should be further explored. We observed that counter-narratives do not change attitudes of convinced antisemites, and opposing their messages can even highlight them and incite antisemites to send more antisemitic posts to back up their initial position. However, critical reactions to antisemitic posts are also seen by others in the respective social media circles and demonstrate that such messages do not go unchallenged. Counter-narratives are more effective with users who are not hardcore antisemites, can initiate critical reflections on stereotypes, and support bystanders in becoming partners in the fight against antisemitism. Counter-narratives can prevent antisemitic norms from being established, and they can give an important voice to anti-antisemites and provide them with resources. Current strategies of counter-narratives, however, are too labor intensive to be implemented on a larger scale. Semi-automated messages that are tailored for certain profiles or messages might help to improve reach and effectiveness.

Cooperation with NGOs combating hate speech against minorities other than Jews makes sense when combating antisemitism from white supremacists who also target other minorities. The same is not true for combating antisemitism from anti-Zionist antisemites who focus their hate on Jews and Israelis. Alliances with organizations combating online support for Islamist terrorism might be more effective here as some of the more radical "anti-Zionist" users endorse terror against Israeli targets and justify terror against Jews as a form of "resistance" against Israel.

[17] World Jewish Congress and Vigo Social Intelligence. "The Rise of Anti-Semitism on Social Media. Summary of 2016," 2017, p. 14, http://vigo.co.il/wp-content/uploads/sites/178/2017/06/AntiSemitismReport2FinP.pdf?x43983 (accessed July 9, 2017).

Annex I: Questionnaire NGO Survey

Questionnaire sent to non-governmental organizations who have been engaged in combatting antisemitism.

1. How do you combat antisemitism on social media? Could you give us an example of the work you are doing on social media, individually or through an organization (which one?)?

2. On which social network platforms and in which languages do you work?

3. Do you have an example from your work in which antisemitism on social media was confronted or reduced successfully?

4. What are your criteria of success?

5. Which main forms of antisemitism and what stereotypes do you observe repeatedly?

6. What groups or individuals have really made a name for themselves in promoting online antisemitism?

7. What kind of collaboration with other organizations do you find helpful? Has the support from governments and multilateral organizations made a difference?

8. What is one of your most difficult challenges in combating antisemitism online? In which area do you need more support?

9. If you had the funding and people to do it, what do you think should be done to combat antisemitism on social media more effectively?

10. Do you think counter narratives can be effective? If so, how? What would be important?

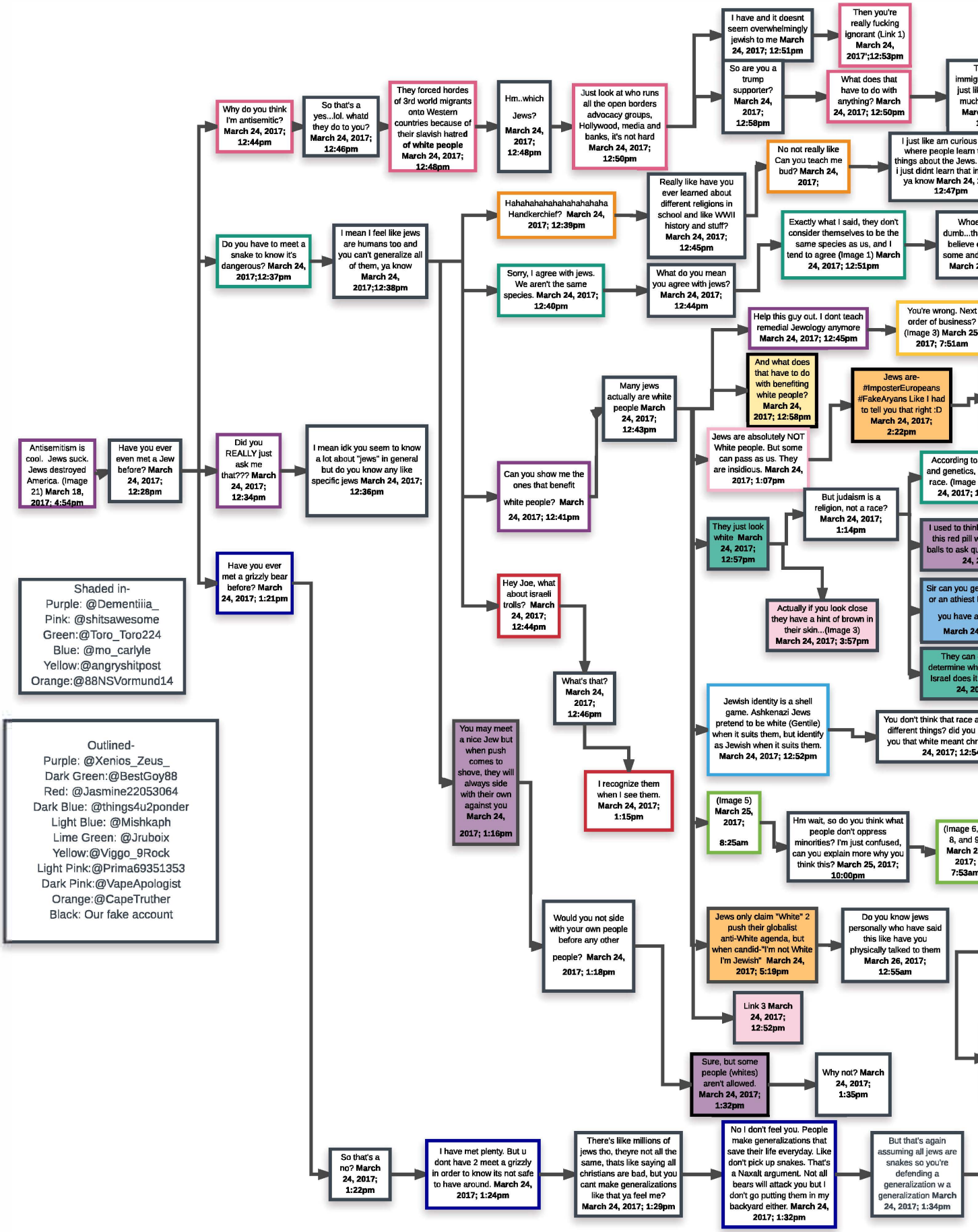
11. Can we publish your response?

___ Yes, entirely: please provide your name (and organization if you wish).

___ Yes, but please keep me anonymous.

12. Do you have any other comments?

Annex II: Example of Attempts to Question An



ntisemitic Messages by People from the Far Right