Antisemitism on Social Media Platforms

Placing the Problem into Perspective

Bossetta, Michael

Published in:
Antisemitism on Social Media

2022

Document Version:
Peer reviewed version (aka post-print)

Antisemitism on Social Media Platforms: Placing the Problem into Perspective

Michael Bossetta (Lund University)

Note: This is a pre-print of a chapter for the book *Antisemitism on Social Media* (Routledge).

Please cite as:

Bossetta, M. (2022). Antisemitism on social media platforms: Placing the problem into perspective. In M. Hübscher & S. von Mering (Eds.), *Antisemitism on social media* (pp. 227-241). Routledge. doi: 10.4324/9781003200499-14

Abstract

This chapter provides a survey of the existing quantitative research on antisemitism and social media. Through doing so, it argues that the sheer quantity of antisemitism on social media is much less than commonly perceived. In addition, the chapter argues that quantitative research often neglects counter-narratives calling out antisemitism, which are an important counterpoint to include in antisemitism research on social media. Then, the chapter discusses how specific components of platform design help explain why antisemitism is likely to surface on some plaftforms but not others. Through this theoretical lens, the chapter compares the design of major social media platforms to online forums. The comparison highlights why memes typically originate in online forums and then are pushed towards more public and mainstream social media platforms. Ultimately, the chapter argues that future research on antisemitism and social media should approach studying antisemitic content across online spaces, with specific attention to the effects of such content on users' potential for radicalization.

To date, research on digital forms and dissemination of antisemitism remains scarce. As evidenced by this volume, academic interest in antisemitism on social media is increasing. The bulk of existing knowledge on the topic mainly stems from two types of research. The first is reports from non-governmental and government bodies, whereas the second is academic investigations into the quantity of antisemitism detected in various online spaces. While the former are valuable in filling empirical gaps, reports by civil society and public sector actors are not peer-reviewed and typically lack the methodological rigor of published scholarly work. Meanwhile, much of the extant scholarship in this area takes the form of conference papers that seek to quantify the amount of antisemitism in online spaces or develop automated tools for doing so.

Unfortunately, measuring the true amount of antisemitism on social media is difficult for four main reasons. First, with the exception of Twitter and Reddit, most social media platforms do not currently offer data on citizens' posts, which scientists need to measure antisemitism in public online spaces. Second, antisemitic content can be shared through private messaging channels (e.g., direct messages on Instagram or Twitter) or social media specifically designed for private messaging (e.g., Facebook Messenger, WhatsApp, or Signal). These private channels are increasingly encrypted, meaning they are unreadable to both the platforms themselves as well as law enforcement. Third, antisemitism can be expressed in disguised or subtle ways, such as through coded phrases or pictures, which makes detecting antisemitism difficult via computational methods. Fourth, the majority of existing research by academics, organizations, and governments focuses on "hate speech" broadly, and rarely do such reports distinguish the amount of antisemitism within overall hate speech content.

Despite these limitations, however, scholars have made progress in detecting and measuring antisemitism on social media. My aim in this chapter is to contextualize the extant quantitative

scholarship in this area through three main arguments. First, I survey the existing research that quantifies the extent of antisemitism on social media to make the argument that its prevalence is much less than commonly perceived. Second, I argue that social media posts challenging antisemitism are an important counterpoint to include in quantitative work. Third, I argue that deconstructing the design components of social media platforms provides a valuable heuristic for understanding the relationship between antisemitism and its likelihood to appear in various online spaces. In particular, I discuss key technological differences between social media platforms and online forums, and explain how the design components of each lend to antisemitism being disseminated across a broad digital ecosystem. Future quantitative work would therefore benefit from theory-driven approaches that consider platform design, in order to inform cross-platform investigations that reveal how antisemitic actors leverage specific platform properties to diffuse their message across the internet.

Diagnosing the Extent of Antisemitism on Social Media

Relative to overall social media traffic, the identified extent of antisemitic content on social media is extremely small. Of course, I do not mean to suggest that online antisemitism is not a problem; rather, the bulk of evidence points to antisemitic content being much less visible on social media than commonly perceived. In this section, I support this claim with a survey of existing empirical research seeking to quantify the extent of antisemitic discourse within various online spaces.

The survey begins in 2016 with a landmark report by the World Jewish Congress (2016). The report is notable, since it diagnoses antisemitism on social media using a cross-platform, cross-country, and mixed-methods approach combining quantitative measurements with qualitative coding. The results of their analysis estimate, conservatively, that approximately 382,000 antisemitic posts were sent on

social media during 2016. Globally, the report estimates that most antisemitic content appeared on Twitter (63%), followed by Facebook (10%) and Instagram (6%).

While the fact that antisemitic content can be identified on social media is troubling, it is important to put these numbers into perspective. On Twitter, the platform where most antisemitic content exists, at least 500 million tweets are sent every day (Stricker 2014). The daily average of images and videos posted to Instagram was 95 million in 2016 (Abutaleb 2016), and 2.5 billion comments were left on public Facebook pages each month in 2015 (Facebook 2015). Across these three platforms, then, the amount of tweets, Instagram posts, and public Facebook comments can be approximated to 250 billion posts per year. Although the World Jewish Congress report did not evaluate every country, the 382,000 antisemitic posts they identified constitutes .00015% of global traffic on these three platforms around the time the report was conducted.

Similarly, the Anti-Defamation League (2018) reports finding an average of 81,400 antisemitic tweets in English on Twitter during 2018. This equates to approximately .0002% of weekly Twitter traffic. These numbers align with a study that examined hate speech on Twitter in the United States during the 2016 election (Siegel et al., 2021) and found through a much more rigorous analysis that antisemitic content rarely rose above .0001% in tweets mentioning Donald Trump (where antisemitism might be expected according to mainstream media reporting).

My point here is to illustrate that non-academic studies that only report raw numbers can lead to misperceptions about the overall scope of antisemitism on social media. The existing academic research fairs better in contextualizing the extent of antisemitism relative to an overall dataset. However, like in the Trump example above, many studies report the percentages of antisemitic

discourse relative to a sampled dataset that, arguably, constitutes oversampling or a case study bias more generally.

Within the context of antisemitism research, case study bias refers to selecting empirical cases where antisemitic discourses are most likely to appear. This can refer to sampling the social media accounts of far-right groups (Weimann & Masri, 2020), media articles by far-right news organizations (Barna & Knap, 2019), or posts in alternative online spaces popular with the far-right (Ridenhour et al., 2020; Zannettou et al., 2020). Other studies sample conversations around events or conspiracies likely to harbor antisemitic content, such as attacks on places of worship (Zelenkauskaite et al., 2021) or mentions of George Soros (Kalmar et al., 2018). While it is important to probe these spaces and events for antisemitism in order to understand its digital manifestations, an emphasis on searching for antisemitism where we know it exists can constitute an oversampling problem. That is, scholars may be overreporting the amount of antisemitism on social media by pre-selecting datasets and cases where it is most likely to surface. More worringly, overattention to obvious spaces for antisemitic content may mask uncovering where it is more subtle and perhaps, more impactful.

Such oversampling is not specific to antisemitism research; most social media platforms that offer data to researchers require data collection based on keywords or specific accounts. This built-in filtering mechanism nudges researchers toward collecting data around specific topics, events, or actors, and the resulting datasets reflect a non-random portion of the overall content circulating on a platform (Tromble, 2019). While oversampling is often unavoidable due to how platforms deliver data, it can still lead to overreporting the existence of digital phenomena if not properly contextualized. Therefore, in reviewing the academic research quantifying antisemitism on social media, I aim to place the

findings into a broader perspective by highlighting the proportions of antisemitism reported in relation to how the studies' datasets were constructed.

As is well-acknowledged in social media scholarship, Twitter receives a disproportionate amount of academic research due to its relatively open data availability in comparison to other platforms. However, since the World Jewish Congress report (2016) suggests that the bulk of antisemitic content is hosted on Twitter, the platform is an appropriate site to measure the quantity of antisemitism. In one of the few peer-reviewed studies of antisemitism on social media, Ozalp et al. (2020) find that *within a dataset of tweets mentioning Jewish keywords and deragotory slurs in the UK*, only .7% of tweets were antagonistic toward Jews. Similarly, Kalmar et al. (2018) find that *in a dataset of tweets mentioning "Soros"*, only .8% of tweets contained the word "Jew." However, the amount of these tweets that could be considered antisemitic was not studied, which is important information to analyze and report.

Research finds that only a fraction of social media posts mentioning Jews are antisemitic. In Jikeli et al.'s (2019) random sample of 11 billion tweets sent during 2018, about 3 million (or .03%) mentioned Jews. Through a manual annotation of 400 randomly selected tweets from this smaller dataset, between 5-6% could be labelled antisemitic with confidence, whereas an additional 7-12% were 'probably antisemitic' (Jikeli et al., 2019, p. 12). Thus, if we take an upper bound of 20% of tweets mentioning Jews to be antisemitic, this comprises .005% of their overall dataset of 11 billion tweets (with the true proportion of antisemitic tweets likely being lower). Wooley and Josef (2019), meanwhile, in their dataset of 5.8 million tweets containing political keywords around the 2018 US election, find that 1.7% contained keywords associated with antisemitism. Their computational analysis suggests that within this 1.7% subset, just under half of tweets (46%) could be classified as derogatory or leaning deragotory toward Jews. Thus, .8% of their electoral tweets may be antisemitic,

a percentage which likely drops significantly when viewed in the totality of non-electoral tweets sent during the same period.

Even in online forums well-known to harbor extreme views, such as 4chan's subcommunity /pol (short for "politically incorrect"), not all posts discussing Jews are antisemitic. One study found that between July 2016 to January 2018, mentions of the word "Jew" comprise between 2-4% of overall posts on /pol, with the derogatory slur "kike" comprising around 1% (Zannettou et al. 2020, p. 4). In qualitatively analyzing a sample of 100 of posts mentioning "Jew" with a binary categorization of hateful/non-hateful, 42% of posts were found to be hateful, with all posts mentioning "kike" considered hateful. Thus, while the authors note that their figures are likely conservative, this equates to less than 2% of posts on /pol being hateful toward Jews with an additional 1% for mentions of kike. Thus, within one of the most hate-filled spaces on the internet, approximately 3% of posts are identified as hateful toward Jews.

Zelenkauskaite et al.'s (2020) data from /pol during February and March 2019 finds similar results. Their results show that terms like "Jew" and "Jewish" ranged between 3.3% - 5.5% in the period during the Christchuch and Pittsburgh shootings at a mosque and synagoguge respectively. Posts with the term "kike" constituted an additional 1.5-2%. Although the authors did not analyze the content of these messages to investigate whether all were hateful, their results suggest that between 4-7.5% of posts on /pol mention Jews or derogatory slurs about them. If findings from the aforementioned studies of both /pol and Twitter apply to their case, less than half of these posts would likely be considered explicitly antisemitic, deragotory, of hateful.

Taken together, the extant research points to antisemitic content being an extremely small percentage of Twitter traffic (fractions of a percent) and in low, single-digit percentages for /pol. While the proportion of antisemitism on /pol is higher than on Twitter, it is important to note that the amount of content generated on /pol is only a fraction of the content generated on Twitter. In one of the largest academic reportings into /pol's activity, scholars estimate that users issue around 150,000 posts per day (Papasavva et al., 2020). That equates to .03% of Twitter's 500 million tweets per day. Thus, although the proportion of antisemitism on /pol is higher than on Twitter, the sheer volume of content on /pol pales in comparison to the daily activity on any major social media platform.

It is perhaps surprising to see the data presented this way, since it suggests a disconnect between the extent of identified antisemitism on social media and the lived experiences of the Jewish community. Jews across Europe rank online antisemitism as one of the most widespread problems facing their communities (EU Agency for Human Rights, 2018, p. 22). There are three potential explanations for this discrepancy between the low proportion of observed antisemitism on social media and the high concerns about online antisemitism expressed by the Jewish community. One likely explanation, and an important contextual factor, is that Jews comprise a minority of the world population (about .2%) and therefore constitute a minority of overall social media users. Thus, the raw numbers of antisemitic posts observed by quantitative studies may be a low proportion of overall traffic but disproportionately affect the Jewish community. Another possible explanation is the "third-person effect," where research shows that people perceive the influence of hate speech on social media as harming others more than themselves (Guo & Johnson, 2020). A third explanation is that, due to limits in data collection, quantitative studies measuring antisemitism cannot observe non-public spaces like direct messages, which may be where higher levels of antisemitic attacks are occurring.

A recent survey by the American Jewish Committee (2020), for example, finds that one-in-five American Jewish adults report being the target of an antisemitic attack or remark either online or on social media. Facebook was the predominant platform for experiencing such an attack or remark (62%), but the survey did not investigate whether these instances occurred through comments on public pages, within private networks, or via direct messages. Data from each of these Facebook channels is currently unavailable to researchers, and therefore the numbers and percentages of antisemitism identified by prior studies likely does not reflect the overall scope of antisemitism online. Thus, in addition to quantitative reporting about its extent, scholars interested in antisemitism on social media should move toward qualitative designs incorporating interviews with those who experience antisemetic attacks, particularly on Facebook. This would help in uncovering where antisemitism takes place on social media and whether its manifestations are more prevalent on channels currently unaccessible for computational research.

Counter-Narratives to Antisemitism on Social Media

In addition to contextualizing the amount of antisemitic content relative to platform traffic, another key component – posts challenging antisemitism – is often overlooked in studies of antisemitism and social media. Studies that report only the raw numbers of antisemitic content may create an implicit assumption that these posts circulate unchallenged, whereas evidence points to the contrary. This section argues that these counter-narratives are valuable from a sociological perspective, as they may stunt the effect of antisemitic content by shaping community norms that refute its presence within a given online community or platform.

According to estimates by the Anti-Defamation League (n.d.), one-in-four individuals hold antisemitic attitudes worldwide. These figures suggest that on aggregate, people not holding antisemic attitudes

outnumber those who do. It is therefore unlikely that antisemitic content on social media will go unchallenged, and existing research supports this notion. In their random sample of tweets mentioning "Jew" and related keywords from 2018, Jikeli et al.'s (2019, p. 12) findings suggest that tweets "calling out antisemitism" outnumbered those that express strong antisemitism. Similarly, the Norwegian Ministry of Local Government and Modernization (2016, p. 23) finds that antisemitic discourses on social media rarely remain unchallenged by others. In a particularly promising finding, Ozalp et al. (2020) find that tweets from Jewish organizations countering antisemitic content have a longer life cycle on the platform than tweets agonistic toward Jews. A longer life cycle signals that counter-narratives to antisemitism receive more engagement – likely in the form of user endorsements – than antisemitic content. Situating antisemitic discourses alongside those that challenge them is important, since social media users can work to correct misinformation (such as conspiracies) or moderate social norms on a platform.

Research on health misinformation, for example, suggests that exposure to user-generated "social corrections" (comments countering false claims) are effective in lowering misperceptions, especially if accompanied by a credible source (Vraga & Bode, 2018). While the study deals with misperceptions around verifiable evidence in the context of disease, it shows how social media users can influence the perceptions of others on a platform. Moreovoer, norms of appropriate behavior within a group are collectively negotiatied. Social media users set the boundaries for appropriate discourse and behavior through social sanctions: rewarding users for following social norms and punishing those who deviate from them (Newman, 2020; Rashidi et al., 2020). Users can positively reinforce posts through reactions (such as "liking" a post on Facebook or Instagram or gifting an "award" on Reddit), leaving positive comments, or sharing a post across their networks to show agreement. By the same token, users can negatively sanction posts by leaving negative reactions (such as an 'angry' reaction on

Facebook or a downvote on Reddit), calling out users in comments, sharing content with added text signaling disapproval, or even reporting content to a community or platform moderator.

In addition to being a form of hate speech, antisemitic posts can violate social norms on a platform along several dimensions, such as spreading false information or being overly aggressive. Each of these factors – hate, false information, and aggression – has been shown to provoke counter-responses from users. In the context of racism and hate, research on Facebook comments to Canadian news finds comments countering racism outnumbered racist comments by a 2:1 margin (Chaudhry & Gruzd, 2020). In a qualitative study, Matamoros-Fernánez (2017, p. 940) highlights how in a racist controversy in Australian football, social media users 'took the lead in denouncing abusive practices' by posting screenshots of offensive content. Aligning with Ozalp et al.'s (2020) findings, Twitter's internal research suggests that during the 2016 US election, retweets calling out disinformation from the Russian Internet Research Agency received eight times the impressions and engagements by ten times as many users than the disinformation tweets themselves (US Senate, 2018). Aggression, meanwhile, has been shown to be a consistent feature that generates replies on both Facebook and news websites (Ziegele et al., 2014). At a more subtle level, even mild disagreement between commenters attracts users' visual attention to Facebook comments relative to comments in agreement (Dutceac Segesten et al., 2020). Although these insights draw from various contexts and sources, they point to a general trend that users actively engage in confronting misleading or offensive content, which can lead to the expression of social sanctions that signal to others that such content is not welcome on a platform.

When sanctions and counter-narratives outweight the volume of antisemitism on a platform, strong social signals are sent to other users about what the community considers appropriate and inappropriate. These sanctions are not likely to inhibit the spreading of antisemitic or racist content

by the vocal minority of users who post it (Chaudhry & Gruzd, 2020); however, they can create the perception that a community of users views such content as violation of social norms. Recent survey research finds that specifically for antisemitic sentiment, perceptions of social norms have significant implications for Jewish prejudice. If individuals perceive that *others* in their community hold positive attitudes toward Jews, then this perception reverses the well-established link between right-wing authoritarianism and out-group prejudice (Górska et al., 2021).

Thus, in cases where challenges to antisemitic posts outnumber the posts themselves, the effect of antisemitic content on the user community is likely attenuated – or even reversed – with sufficient exposure to counter-narratives. Experimental designs are best suited to test this hypothesis, and future research should measure the effect of antisemitic posts on users' attitudes toward Jews with and without exposure to counter-narratives. Importantly, however, this argument about the efficacy of counter-narratives primarily applies to the perceptions of the user collective. Counter-narratives may do little to attenuate the effects of a targeted, antisemitic attack to an individual user. Still, counter-narratives and their potential to shape group norms are an important component of social media research, and like antisemitic attacks, will depend on the technical design of an online space.

Antisemitism on Social Media versus Online Forums

Within a given online space, the prevalence of antisemitic content, its ability to be targeted, and the opportunity for counter-narratives are influenced by the technological design of that space. Since the design of a platform shapes digital communication (Bossetta 2018), certain elements of platform design will directly impact the type of antisemitic content produced, how and where it is shared, and between whom. In this final section, I outline four key features of platform design and how they relate to the production, reception, and spread of antisemitic content across social media.

Supported media refers to the type of media that a platform allows to be uploaded and, in many cases, actively promotes. For example, YouTube is designed primarily to support video, Instagram is built mostly for images, and Facebook is flexible in supporting text, images, and video. When it comes to antisemitic propaganda, the most common media format is images (World Jewish Congress, 2016, p. 169) and there are two plausible explanations for this – one cultural and the other technical. First and culturally, images are powerful, not tied to a specific language, and they can depict symbols that are packed with historical meaning. Second and technically, images are easier to produce than video, and they also take up less storage memory on a device. This means that images are often easier and cheaper to produce than video, and they can be shared more quickly between users since they require less bandwidth from an internet or mobile network. As an added security measure for users who spread antisemitic content, images can be more difficult for security systems to detect, in comparison with offensive keywords often used to identify antisemitism in research applications.

However, when it comes to spreading content on social media, *sharing features* are a key element of platform design that influences the visibility of content on a platform. Sharing can occur in three ways: sharing privately between individuals or groups (such as on WhatsApp or Facebook Messenger), sharing original content to a social network (such as posting a photo on Facebook or Instagram), or sharing someone else's content to a social network (such as retweeting a news article on Twitter or sharing a friend's post on Facebook). For antisemitic content to be widely seen, it typically needs to be amplified by many accounts in public spaces. Most social media – with the exception of Facebook, Twitter, and LinkedIn – lack a feature to publicly share content from another account.

This is where *user connections* come into play: the accounts that a user connects with on social media and the rules behind how these connections are formed. "Social" media is about connecting people, and a user's online social network is influenced by several factors. Some platforms – such as Instagram, Snapchat, Facebook, and WhatsApp – encourage connections between close friends. Others like YouTube or TikTok are more about connecting users to entertaining videos, and therefore serve a different purpose than relationship-building with friends in real-life. A key element of user connections that is directly relevant for antisemitism is whether accounts are allowed, technically, to be anonymous. Platforms that encourage connections between real-life friends generally require accounts to be verified with a phone number, which helps authenticate the identity of a user to an account. Other, more public platforms allow users to remain anonymous and have less strict identity verification rules. Research shows that discussions on platforms that allow anonymity tend to be less civil than on Facebook, where one's identity is tied to an account (Halpern and Gibbs 2013). Moreover, platforms that allow anonymity are more likely to have antisemitic content, since the spread of online hate speech can be punishable by law in some countries, like Germany.

Whereas anonymity may shield an individual user from legal accountability, social media platforms are increasingly under pressure to adjust their *content moderation policies,* which refers to the rules and enforcement of rules by social media companies about what type of content to allow on their platforms. All social media moderate content (Gillespie 2018), but the enforcement of those policies can vary based on the size of the platforms. Bigger, more established social media like Facebook and YouTube have more resources to dedicate to removing content, so antisemitic content is less likely to stay visible on those platforms for very long. European Commission reports show that both Facebook and YouTube remove about 80% of reported hate speech content in Europe, compared to Twitter's 40% (Reynards, 2020). While all of these platforms have explicit content moderation policies banning

hate speech, their ability to enforce these policies likely boils down to the amount of resources that each platform dedicates to enforcing them.

My aim here has been to briefly illustrate how four elements of platform design – supported media, sharing features, user connections, and content moderation policies – shape the type, visibility, and reach of all content on social media, including that which is antisemitic. From this analysis, antisemitic content is likely to spread on platforms that: support images, have an easily accessible sharing feature (such as reweeting), allow users to be anonymous, and have less restrictive or enforceable content moderation policies. This is likely why Twitter has been detected as the mainstream social media platform hosting the most antisemitic content (World Jewish Congress, 2016), since it fulfills all of these criteria.

However, there are significant differences between the design of traditionally understood *social media platforms* like Instagram, Facebook, or Twitter and *online forums* like 4chan or Reddit. From a design perspective, online forums typically support only images, text, and hyperlinks; allow for anonymity; and do not contain sharing features within the site. Moreover, they have almost no content moderation policies, except for those that human moderators – or "mods" – decide to impose on any specific subcommunity. Thus, these forums can allow content that expresses opinion that are viewed as extreme, hateful, and racist by the majority of society. Within online forums, anonymous individuals meet to exchange opinions on these topics, create and share images that perpetuate antisemitic tropes (Zannettou et al., 2020), and discuss how to convert susceptible individuals to share their extremist worldview. By becoming involved in these online communities, users can engage with anonymous individuals that often have the explicit intent to manipulate and radicalize others toward their cause.

Crucially, a key design difference between most social media platforms and online forums is how users access content. On social media, algorithms curate and tailor personalized content recommendations for users, often according to what one's *user connections* have redistributed through *sharing features*. By contrast, users of online forums must actively navigate to specific subcommunities (such as 4chan's /pol or subreddits on Reddit), which are usually divided by topics. Therefore, users largely self-select into topical communities that can become congregation spaces for the like-minded. Rossini's (2020) research shows that intolerant discourse is more likely to occur in Facebook threads where users express ideological agreement, a finding that suggests hate speech is more likely to occur in homophilous networks (i.e., communities of the like-minded) compared to diverse networks. Since online forums typically lack the algorithmic curation that expose users to diverse opinions (Bakshy et al., 2015), specific subcommunities within these forums – like /pol – can foster environments of like-minded users who can access lightly moderated and algorithmically unfiltered content. Moreover, the lack of opinion diversity within these spaces makes counter-narratives less likely, and therefore the promotion of antisemitic views is largely rewarded as pro-social norm, rather than negatively sanctioned as inappropriate.

Some of these subcommunities harbor discussions typically associated with the "alt-right," such as white supremacy, Islamophobia, and antisemitic views. Members who participate in these discussions perceive of themselves as a reactionary counter-culture against political correctness, which they view as stemming from the political left, mainstream media, globalists, and feminists. They therefore turn to anonymous online forums to vent their frustration, discuss news, and more worryingly, coordinate on how to radicalize susceptible individuals to share their worldview. Although there is not much concrete evidence on who is likely to be radicalized on these forums, the general impression is that they are young, white men who lack a rich social network in real life, especially in terms of dating

women (Marwick and Lewis, 2018). Most often, users in these forums seek to radicalize others through the creation and sharing of "memes," images that are designed to be humorous and packed with ideological and cultural symbols. Even with modern deep learning techniques, detecting antisemitism in such images is difficult to automate due to the subtleties of sarcasm and irony often carried in memes (Chandra et al., 2021).

Nevertheless, research has been able to identify how antisemitic memes flow *from* online forums *onto* mainstream social media sites. Zannettou et al (2020) found evidence that 4chan's /pol subcommunity, as well as Reddit's The_Donald (a subcommunity built around supporting of President Trump), were effective in promoting antisemitic memes on Twitter. This two-step flow of meme dissemination likely signifies a recruitment effort: since the user connections within an online forum like /pol are typically like-minded, forum members need to reach out to more diverse user bases on mainstream social media to grow their movement. Technically, since both mainstream social media and online forums support images, forum members can develop and share memes within their homogenous networks before attempting to disseminate them more widely on mainstream social media.

For an episode of the Social Media and Politics podcast, I interviewed one of the original moderators of 'The_Donald' (Bossetta, 2019), a subreddit linked to the spreading of antisemitic memes (Zannettou et al., 2020). Although our discussion focused on the subreddit's political motivations (and not antisemitism explicitly), the moderator mentioned three key properties of a successful meme, why they are impactful in recruiting new members to their worldview, and how memes work as part of a broader process of ideological recruitment.

According to him, the three key components of a successful meme are *simplicity, humor,* and *exposure*. Simplicity – i.e., a single picture rather than a video or long text post – is vital to capture users' attention in a highly saturated media environment. The moderator mentioned "fun," or humor, as an important criteria for a successful meme because it incentivizes group association: "you want to be the fun person, and you want to be with the fun person." Memes that achieve humor aim to tap into group psychology mechanisms that provoke emulation ("to be the fun person") and pro-social association ("to be with the fun person").

Importantly, the third aspect – exposure – is crucial to understanding that memes operate as the first tranche of a broader ideological recruitment process. That is, memes are used to desentize users to certain ideas, generate interest in learning more about them, and ultimately aim to have users' redistribute the meme across their own social networks. Repeated exposure to memes help to desensitize users to controversial ideas and generate interest in exploring them further, as noted by the moderator: "If you have a topic that gets tons and tons of memes, people casually browsing the internet keep running into the same idea over and over. Eventually, they might look into it."

After a user follows up on a meme's subject matter through reading articles or posts in online forums about it, the goal of meme makers is to have converted users distribute memes on mainstream platforms via their sharing features. According to the moderator:

> "A meme that's funny but also has that little redpill, or that little bit of truth…if somebody shares that on their Facebook or Twitter, people are going to ask about or comment on it. It's gonna generate conversation, then people start posting articles. And, it kind of solidifies your understanding of it and you might share out from there."

Thus, the flow of antisemitic memes that Zannettou et al. (2020) identify – from online forums onto mainstream media – can be partly explained through the lens of platform design. Since online forums

have light *content moderation policies* and homogenous *user connections*, antisemitic discourse can thrive largely unchallenged in these spaces. Although lower in technical sophistication than social media platforms, online forums *support images*; however, they lack *sharing features* necessary to distribute them widely. In order to recruit new members and grow their ideological movement, forum users need to "push" antisemitic memes onto mainstream social media channels, with the ultimate aim of turning converts into dissemination nodes that share memes with their own personal networks. To better investigate how antisemitism spreads across the digital ecosystem in this manner, future research would strongly benefit from developing theory-driven approaches that use platform design to motivate empirical case selection, rather than seeking to quantify antisemitic content on any given platform.

Conclusion

This chapter aimed to place antisemitism on social media into perspective on three fronts. Through a survey of the existing quantitative research on antisemitism and social media, I first argued that the amount of antisemitism on social media is extremely small relative to platform traffic. Second, existing quantitive studies often do not consider counter-narratives to antisemitism, which I argue is an important component in providing a full perspective of how antisemitism is received (and refuted) by social media users. That said, antisemitic content continues to circulate across the internet, and my third argument has been that platform design offers theoretical insight into why and how antisemitism surfaces across online spaces.

Ultimately, this chapter concludes by arguing that the sheer quantity of antisemitism on social media is neither a primary cause for concern nor a pressing task for research. It is not the sheer quantity, but rather the potential for antisemitic content to radicalize, that makes it an important issue for

democratic societies. While limiting the visibility of antisemitism online is an important step to detoxify users' online information environment, much more effort needs to be placed on understanding who is susceptible to being radicalized by such content and why. Only then will the role of online antisemitism be fully placed into perspective.

References

Abutaleb, Y. (2016). Instagram's User Base Grows to More Than 500 million. *Reuters,* [online]. Available at: https://www.reuters.com/article/us-facebook-instagram-users/instagrams-user-base-grows-to-more-than-500-million-idUSKCN0Z71LN [Accessed 9 July 2021].

American Jewish Committee (2020). *The State of Antisemitism in America 2020: AJC's Survey of American Jews.* [online]. Available at: https://www.ajc.org/AntisemitismReport2020/Survey-of-American-Jews [Accessed 9 July 2021].

Anti-Defamation League (2018): *Quantifying Hate: A Year of Antisemitism on Twitter.* [online]. Available at: https://www.adl.org/resources/reports/quantifying-hate-a-year-of-antisemitism-on-twitter#major-findings [Accessed 9 July 2021].

Bakshy, E., Messing, S. and Adamic, L.A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science*, 348(6239), pp.1130-1132.

Barna, I. and Knap, Á. (2019). Antisemitism in Contemporary Hungary: Exploring Topics of Antisemitism in the Far-right Media using Natural Language Processing. *Theo Web*, 18(1), pp.75-92.

Bossetta, M. (2019). *Pro-Trump social networks: The Donald on Reddit and TheDonald.win.* [podcast]. Social Media and Politics. Available at: https://socialmediaandpolitics.org/pro-trump-social-networks-the-donald-reddit-thedonald-win/ [Accessed 9 July 2020].

Bossetta, M. (2018): The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 US Election. *Journalism & Mass Communication Quarterly*, 95(2), pp. 471-496.

Chandra, M., Pailla, D., Bhatia, H., Sanchawala, A., Gupta, M., Shrivastava, M. and Kumaraguru, P. (2021). "Subverting the Jewtocracy": Online Antisemitism Detection Using Multimodal Deep Learning. In: *Proceedings of the 13th ACM Web Science Conference.* Virtual: ACM, pp. 148-157.

Chaudhry, I. and Gruzd, A. (2020). Expressing and Challenging Racist Discourse on Facebook: How Social Media Weaken the "Spiral of Silence" Theory. *Policy & Internet*, 12(1), pp.88-108.

Dutceac Segesten, A., Bossetta, M., Holmberg, N. and Niehorster, D. (2020). The Cueing Power of Comments on Social Media: How Disagreement in Facebook Comments Affects User Engagement with News. *Information, Communication & Society*. [online]. Available at:

https://www.tandfonline.com/doi/full/10.1080/1369118X.2020.1850836 [Accessed 9 July 2021].

European Union Agency for Fundamental Rights (2018). *Experiences and Perceptions of Antisemitism: Second Survey on Discrimination and Hate Crime against Jews in the EU*. [online]. Available at: https://www.jfm.se/wp-content/uploads/2018/12/fra-2018-experiences-and-perceptions-of-antisemitism-survey_en.pdf [Accessed 9 July 2021].

Facebook (2015): *New Tools for Managing Communication on Your Page.* [Blog] Facebook for Business. Available at: https://www.facebook.com/business/news/new-tools-for-managing-communication-on-your-page?__mref=message_bubble [Accessed 9 July 2021].

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Górska, P., Stefaniak, A., Lipowska, K., Malinowska, K., Skrodzka, M. and Marchlewska, M. (2021). Authoritarians Go with the Flow: Social Norms Moderate the Link between Right-Wing Authoritarianism and Outgroup-Directed Attitudes. *Political Psychology.* [online]. Available at: https://onlinelibrary.wiley.com/doi/epdf/10.1111/pops.12744 [Accessed 9 July 2021].

Guo, L., & Johnson, B. G. (2020). Third-Person Effect and Hate Speech Censorship on Facebook. *Social Media + Society,* 6(2), [online]. Available at: https://journals.sagepub.com/doi/full/10.1177/2056305120923003 [Accessed 9 July 2021].

Halpern, D. and Gibbs, J., 2013. Social Media as a Catalyst for Online Deliberation? Exploring the Affordances of Facebook and YouTube for Political Expression. *Computers in Human* Behavior, 29(3), pp.1159-1168.

Jikeli, G., Cavar, D. and Miehling, D. (2019). Annotating Antisemitic Online Content. Towards an Applicable Definition of Antisemitism. [online]. [Preprint]. Available at: https://arxiv.org/abs/1910.01214 [Accessed 9 July 2021].

Kalmar, I., Stevens, C. and Worby, N. (2018). Twitter, Gab, and Racism: The Case of the Soros Myth. In: *Proceedings of the 9th International Conference on Social Media and Society.* Copenhagen: ACM, (pp. 330-334).

Marwick, A. and Lewis, R. (2018): *Media Manipulation and Disinformation Online.* [online]. Available at: https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf [Accessed 7 July 2021].

Norwegian Ministry of Local Government and Modernization (2016): *Action Plan against Antisemitism 2016-2020*. [online]. Available at: https://www.regjeringen.no/contentassets/dd258c081e6048e2ad0cac9617abf778/action-plan-against-antisemitism.pdf [Accessed 7 July 2021].

Newman, D. 2020. *Sociology: Exploring the architecture of everyday life.* London: Sage.

Ozalp, S., Matthew, L., Williams, P. B., Han L, and Mohamed M. (2020). Antisemitism on Twitter: Collective Efficacy and the Role of Community Organizations in Challenging Online Hate Speech. *Social Media + Society,* 6(2). [online]. Available at: https://journals.sagepub.com/doi/10.1177/2056305120916850 [Accessed 9 July 2021].

Papasavva, A., Zannettou, S., De Cristofaro, E., Stringhini, G. and Blackburn, J. (2020). Raiders of the Lost Kek: 3.5 Years of Augmented 4Chan posts from the politically incorrect board. In: *Proceedings of the International AAAI Conference on Web and Social Media.* Atlanta: AAAI Press, pp. 885-894.

Rashidi, Y., Kapadia, A., Nippert-Eng, C. and Su, N.M. (2020). "It's Easier than Causing Confrontation": Sanctioning Strategies to Maintain Social Norms and Privacy on Social Media. In: *Proceedings of the ACM on Human-Computer Interaction*. Minneanapolis/Virtual: ACM, pp.1-25.

Ridenhour, M., Bagavathi, A., Raisi, E. and Krishnan, S. (2020). Detecting Online Hate Speech: Approaches Using Weak Supervision and Network Embedding Models. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation.* Virtual: Springer, pp. 202-212.

Reynards, D. (2020). *Countering Illegal Hate Speech Online: 5th Evaluation of the Code of Conduct.* [online]. Available at: https://ec.europa.eu/info/sites/default/files/codeofconduct_2020_factsheet_12.pdf [Accessed 7 July 2021].

Rossini, P. (2020). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research.* [online]. Available at: https://journals.sagepub.com/doi/full/10.1177/0093650220921314 [Accessed 9 July 2021].

Siegel, A.A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J. and Tucker, J.A. (2021). Trumping Hate on Twitter? Online Hate Speech in the 2016 US Election Campaign and its Aftermath. *Quarterly Journal of Political Science*, 16(1), pp.71-104.

Stricker, G. (2014). The 2014 #YearOnTwitter. [Blog]. Twitter Blog. Available at: https://blog.twitter.com/official/en_us/a/2014/the-2014-yearontwitter.html [Accessed 9 July 2021].

Tromble, R. (2019). In Search of Meaning: Why We Still Don't Know What Digital Data Represent. *Journal of Digital Social Research*, 1(1), pp.17-24.

US Senate (2018). *Questions for the Record, Senate Select Committee on Intelligence, Hearing on Social Media Influence in the 2016 US Election.* [online]. Available at: https://www.intelligence.senate.gov/sites/default/files/documents/Twitter%20Response%20to%20Committee%20QFRs.pdf [Accessed 7 July 2021].

Vraga, E.K. and Bode, L. (2018). I do not Believe You: How Providing a Source Corrects Health Misperceptions across Social Media Platforms. *Information, Communication & Society*, *21*(10), pp.1337-1353.

Weimann, G. and Masri, N. (2020). Research Note: Spreading Hate on TikTok. *Studies in Conflict & Terrorism*. [online]. Available at: https://www.tandfonline.com/doi/full/10.1080/1057610X.2020.1780027 [Accessed 9 July 2021].

Woolley, S., and Joseff, K. (2019). *Jewish Americans: Computational Propaganda in the United States.* [online]. Available at: https://www.iftf.org/fileadmin/user_upload/downloads/ourwork/IFTF_JewishAmerican_comp.prop_W_05.07.19.pdf [Accessed 7 July 2021].

World Jewish Congress (2016): *The Rise of Antisemitism on Social Media: Summary of 2016*. [online]. Available at: https://www.worldjewishcongress.org/download/RVsVZzRXTaZwO41YbzlWwg?utm_source=WJC+Mailing+Lists&utm_campaign=78bfed156d-EMAIL_CAMPAIGN_2018_02_08&utm_medium=email&utm_term=0_04292c525e-78bfed156d-318920277 [Accessed 7 July 2021].

Zannettou, S., Finkelstein, J., Bradlyn, B. and Blackburn, J. (2020). A Quantitative Approach to Understanding Online Sntisemitism. In: *Proceedings of the International AAAI Conference on Web and Social Media.* Atlanta: AAAI Press, pp. 786-797.

Zelenkauskaite, A., Toivanen, P., Huhtamäki, J. and Valaskivi, K. (2021). Shades of Hatred Online: 4chan Duplicate Circulation Surge during Hybrid Media Events. *First Monday*, 26(1). Available at: https://firstmonday.org/ojs/index.php/fm/article/view/11075/10029 [Accessed 9 July 2021].

Ziegele, M., Breiner, T. and Quiring, O. (2014). What Creates Interactivity in Online News Discussions? An Exploratory Analysis of Discussion Factors in User Comments on News Items. *Journal of Communication*, 64(6), pp.1111-1138.