Political Behavior (2024) 46:1747–1769
https://doi.org/10.1007/s11109-023-09892-9

ORIGINAL PAPER

From Anti-Muslim to Anti-Jewish: Target Substitution on Fringe Social Media Platforms and the Persistence of Online and Offline Hate

William Hobbs[1] · Nazita Lajevardi[2] · Xinyi Li[1] · Caleb Lucas[3]

Accepted: 1 August 2023 / Published online: 7 September 2023

Abstract

The 2016 presidential campaign saw high levels of anti-Muslim online and offline hate. But, by the August 2017 'Unite the Right' rally, anti-Muslim discourse and hate crimes had partly receded, despite the group remaining politically salient and despite a sharp increase in White 'nationalist' activity targeting another religious minority, Jews. Was this by chance? Because we might expect White nationalist activity to increase hate against all groups, the counter-intuitive decline in anti-Muslim hate could have been coincidental. We argue instead that those shifts in animus toward Muslims and Jews should be considered in tandem, and that these over-time patterns of hate reflected different manifestations of elevated and constant religious ethnocentrism, especially among far-right extremists. Using data on fringe and mainstream social media sites and hate crime databases, we present two core sets of findings. First, increased anti-Jewish speech was partly driven by the same far-right communities and extremists who previously promoted anti-Muslim speech. Moreover, combined anti-Muslim and anti-Jewish rhetoric in fringe far-right social media over this period was sustained at a high and largely constant level, seeing shifts primarily in the *targets* of hate speech. Second, similar patterns manifest offline: hate crimes were more strongly associated with *which* group was targeted by hate speech, but not the overall prevalence of hate speech. Together, this study demonstrates a robust link between the dislike toward Muslims and dislike toward Jews, and how fringe groups organize the dissemination of hate.

We are grateful to the editors and the anonymous reviewers for their invaluable feedback. We also thank Marisa Abrajano, Per Adman, Abbas Barzegar, Taylor Carlson, Vicky Fouka, James Fowler, John Kuk, Zachary Steinert-Threlkeld, and Jakana Thomas for their helpful comments and suggestions as we developed the manuscript. Finally, we thank Sophia Lada and Marissa Rivera for their research assistance. All remaining errors are our own.

Extended author information available on the last page of the article

Recent years have seen an uptick in U.S. hate crimes, and extremist rhetoric in day-to-day American politics become more explicit (Giani & Méon, 2019; Mathew et al., 2019; Müller & Schwarz, 2023, 2021; Nithyanand et al., 2017; Siegel et al., 2021). Anti-Muslim hate, in particular, rose throughout the 2016 presidential campaign, with numerous studies documenting that anti-Muslim animus emerged as among the most important determinants of American political attitudes and presidential vote choice (Dunwoody & McFarland, 2018; Jardina & Stephens-Dougan, 2021; Lajevardi & Abrajano, 2019; Oskooii et al., 2021; Tesler, 2022). This rise in anti-Muslim hate during the presidential campaign was followed by an escalation in White 'nationalist' activity. During the Trump presidency, extremists and conspiracy theorists established themselves on fringe social media platforms, while others reportedly moved to these sites after January 6, 2021.[1] By August 2017, when the far-right "Unite the Right" rally took place, the term *nationalist* had become synonymous with "White supremacist," for many.

In this paper, we study how the dynamics of hate in these fringe communities can differ from mainstream spaces, and the implications of those dynamics for offline hate crimes. We argue that while hate targeted Muslims specifically in 2015 and 2016, by mid-2017, hate speech and hate crimes in *fringe, extremist* social networking communities shifted targets from one minoritized religious group, Muslims, to another, Jews. That is, although Muslims were explicitly targeted early in the 2016 presidential campaign, the effects of this rhetoric on increasingly explicit expressions of hate were not constrained to only be anti-Muslim in the long-run. Instead, the effects of increasing hate targeting Muslims may have generalized (Kalkan et al., 2009), or *reverted* to others, especially Jews, within fringe communities increasingly organizing around both White and Christian nationalism and the belief that America is a "Christian nation" (Baker et al., 2020; Thompson, 2021).

We focus on fringe social media sites, and compare our findings to studies of hate on more mainstream platforms, because fringe communities draw in many extremists, including neo-Nazis, and such users might sustain and escalate one another's hate (Walther, 2022). With heightened and sustained animus—and the perception of *all* religious minorities as foreign threats—the dynamics of hate in fringe communities could differ significantly from mainstream ones and in particular ways. For instance, these communities could experience rapid and relatively contiguous shifts in the targeting of outgroups. In mainstream media, there are often many competing news stories with a small fraction of coverage dedicated to any one story over a long period of time (Boydstun, 2013). Although mainstream public opinion is also influenced by selective reporting and stereotyped accounts of events (Zaller, 1992), we expect topics related to outgroup hate to persist longer in fringe settings than in the mainstream media, with fewer periods without a highly salient outgroup.

For this context, we propose the concept of target substitution, where users of fringe sites might substitute their targets of hate. Target substitution—or constant but

[1] Gab, for example, gained over 500,000 users in the days after the riot (Stimson, 2021). Simultaneously, offline, far-right rallies became increasingly commonplace, influenced, in part, by Trump's 2016 presidential campaign rhetoric (Newman et al., 2021).

redirected hate—could be driven by a variety of mechanisms including shifts in the salience of competing considerations (see Zaller 1992, but also see Wlezien 2023), changes in the salience or classification of out-groups, such as deviance theory in sociology (Erikson, 1966), changes in in-group boundaries (Fouka & Tabellini, 2022), or *relative* perceived threat through 'group reference dependence' (Cikara et al., 2022). We expect such mechanisms to be magnified and relatively constant on fringe sites, venues that are most likely cases for target substitution, given dedicated amplification of ethnocentric and xenophobic topics and a greater fraction of users with concordant attitudes. Because of this, 'mainstream' increases in inflammatory rhetoric that lead to a greater user base on fringe sites could provide larger pockets of coordinated extremists, which could sustain and direct hate to new targets well after the original gateway rhetoric declines.

Our time period for this analysis is on hate directed toward U.S. Jews and Muslims in the years and months around the time of the Charlottesville 'Unite the Right' rally in August 2017. We draw on data from fringe and mainstream online discussions (4chan, Gab, Reddit) and hate crime databases (ADL, CAIR, FBI) to examine the sustainability of hate against Jews on fringe platforms and its manifestation offline after the salience of Muslims declined.

Two core findings emerge from our analyses. First, anti-Muslim hate speech on fringe social media sites declined in the month preceding the "Unite the Right" rally, only to be supplanted by anti-Jewish rhetoric the next month. Analyses of user-level posts suggest that the same extremists previously promoting anti-Muslim and anti-Arab speech transitioned to disseminating anti-Jewish rhetoric. Second, we observe nearly contemporaneous declines in offline hate crime targets during periods of declining online hate speech against Muslims and rising hate speech against Jews, both prior to and following the 'Unite the Right' rally, as well as on a week-to-week basis in 2017 and 2018. Strikingly, no such associations were found for *combined* mentions of Jews and Muslims on these sites, suggesting the possible substitution of hate towards one target for another within these fringe communities. Thus, relative targeting within fringe communities served as a stronger indicator of offline hate than the change in expressed hate against all groups on these sites.

Together, this research enhances our empirical understanding of intergroup attitudes by demonstrating a robust link between hate directed at Muslims and Jews. It also highlights the significant role that fringe platforms could continue to play in mobilizing extremists, providing a space where hate can easily shift from previous targets to new ones. Finally, and consequentially, over-time patterns of hate crimes *experienced* by Muslims and Jews offline mirrored the hate disseminated within these fringe communities. These findings illustrate the need for scholars to broaden the scope of their analyses when assessing hate, and that group-by-group analyses can miss important trends in hate speech and hate crimes. In fringe communities—which we might expect to account for a disproportionate fraction of hate targeting minoritized groups—hate can persist with or without mainstream influences.

Theorizing Opposition to Muslims and Jews

Scholars have long recognized that those rejecting one outgroup often reject others too. However, the connection between opposition towards Muslims and other groups remains contested. As Allport (1954) once posited, "If a person is anti-Jewish, he is likely to be anti-Catholic, anti-Negro, anti any outgroup" (p. 68). This suggests that hostility towards outgroups like Jews and Muslims—considered unequal, threatening, and culturally deviant—might be sub-aspects of a more general outgroup devaluation (Meuleman et al., 2019; Zick et al., 2008a).

Several theories suggest that attitudes towards Muslims particularly associate with attitudes towards other cultural out-groups (e.g. Mason et al., 2021; Kam & Kinder, 2012; Stangor et al., 1991; Zick et al., 2008b). This ethnocentric framework divides the world into 'friends' or 'foes' (Kam & Kinder, 2012). Scholars who have applied this theory to anti-Muslim attitudes argue that these often align with sentiments toward other minoritized out-groups (Kalkan et al., 2009), as negative evaluations along racial, cultural, and religious lines typically predict opposition to Muslims. Recent scholarship indicates that anti-Muslim attitudes and xenophobia strongly mediate Christian nationalism's effect on Trump vote intentions (Baker et al., 2020).

However, exclusively focusing on prejudice's generalized nature can also hamper our understanding (Meuleman et al., 2019). Some U.S. studies have shown that anti-Muslim stereotypes independently and robustly affect policy and candidate preferences, even after controlling for a host of social group attitudes (Jardina & Stephens-Dougan, 2021; Lajevardi & Abrajano, 2019; Saleem et al., 2017). This highlights the importance of distinguishing anti-Muslim animosity, given the rising discrimination against U.S. Muslims in politics and in day-to-day social contexts over the past two and a half decades (Hobbs & Lajevardi, 2019; Lajevardi, 2020; Oskooii et al., 2021).

These approaches need not be mutually exclusive. Islamophobia and anti-Semitism can be understood as sub-aspects of xenophobia, with attitudes towards Muslims and Jews rooted in the perception of them being culturally, politically, and theologically distinct (Penning, 2009). Meuleman et al. (2019) find a common denominator of generalized prejudice in attitudes towards immigrants, Muslims, Jews, and sexual minorities. Despite these attitudes possessing group-specific components rooted in different levels of realistic, socioeconomic, symbolic, or cultural threat, they still link to perceptions of out-group threat. Despite differing reasons for fear and anger towards Muslims and Jews, both can boost support for alt-right ideology, fostering a perception of threatened lifestyles and unprotected interests (Isom et al., 2021). Minority threat theory further suggests that as minority groups increase, those with power and status will seek social control over them (Blalock, 1967; Olzak, 1994), a concept corroborated by abundant research linking minority threat to white interest in alt-right ideology and media (e.g., Kyler & Charron-Chénier, 2022; Isom et al., 2021; McVeigh, 2009).

Linked Attitudes Among Extremists and Target Substitution

Public opinion research typically studies outgroup attitudes in representative U.S. population samples, yet Islamophobia and anti-Semitism may be more correlated among extremists than the general public. The distinction between mainstream and extremist perspectives is crucial for our analyses; we focus on how the dynamics of hate on fringe social media sites might differ from over-time patterns of hate more generally. Gerteis and Rotem (2022)'s research using Latent Class Analysis unveiled a correlation between anti-Muslim and anti-Jewish attitudes among a subgroup of White Americans, finding that the two sets of group attitudes share a "cultural logic as connected forms of ethno-religious boundary-making."

To illustrate this work and to make clear our mainstream versus extremist distinction, we turn to the same 2014 Mosaic dataset[2] employed by Gerteis and Rotem (2022), and visualize support for negative stereotypes of Jews and Muslims (see Fig. 1). Our review confirms their findings and underscores two takeaways: (1) A larger segment of White, Non-Hispanic respondents (excluding Jews or Muslims) hold anti-Muslim attitudes than anti-Jewish attitudes, and (2) A distinct group, high in anti-Jewish and anti-Muslim attitudes—"Extremists high in Religious Ethnocentrism"—are clustered and distinct from the general population. As Fig. 1 demonstrates, among this subset, animus towards Muslims and Jews are more related (if not conflated) than the public at large.

We introduce this figure for two reasons: (1) for face validity reasons to demonstrate the existence of individuals with high anti-Semitic and anti-Muslim attitudes, and (2) to propose that this group is likely to frequent fringe social media sites, and thus be amenable to substituting one religious out-group for another when site discourse changes. Among this subset, animosity toward Jews and Muslims is particularly correlated, possibly manifesting in explicit hate speech and crimes, and making them a vital focus of study.

Our target substitution effect hypothesis resembles aspects of attitude change in public opinion studies (Zaller, 1992), where attitudes are influenced by the salience of competing considerations. Though surveys suggest out-group animus generalizes across groups, the same latent animosity can be expressed differently over time in hate speech and hate crimes. We posit that extremists harbor similarly negative (latent) attitudes toward both groups, with their targeting influenced by each group's salience.

We also explore the possibility that salience for all groups can increase or decrease, reflecting varying ethnocentrism levels, *or* that fringe communities operate under a hate 'budget', indicating constant ethnocentrism levels and more dynamic target substitution. And, although we believe that many users will have latent anti-Semitic attitudes as implied by our presentation of survey attitudes in Fig. 1, we leave open the possibility that some users on fringe sites who would not have previously expressed anti-Semitic views will do so after increased exposure

[2] https://www.thearda.com/data-archive?fid=BAM14.

to and interactions with neo-Nazis on the platforms (in the figure, moving from the bottom-right to top-right over time).

Case and Tests

Our study investigates the implications of religious ethnocentrism among extremists on hate speech and hate crimes over time. We question whether increased anti-Semitic online and offline attacks coincide with increased anti-Muslim ones, or if extremists might coordinate these attacks separately.

Immediately after the 2016 election, Jews and Muslims faced heightened threats and security concerns.[3] Concurrently, fringe social media sites attracted numerous new members during the Trump presidency and even moreso after the 2021 Capitol riot. Despite geographical distance, these minimally moderated sites have facilitated closer contact between extremists (e.g. Hafez & Mullins, 2015; Hawdon et al., 2019; Bloom et al., 2019). While some social media sites cracked down on inflammatory rhetoric, others—like 4chan, 8chan, and Gab—continued to provide platforms for speech ingrained with racism, sexism, religious bigotry, anti-LGBTQ+ attitudes, and anti-immigrant animus.[4]

Our study centers the "Unite the Right" rally as a watershed event for White supremacist attitudes in the U.S. The Charlottesville in August 2017 has been called a 'coming-out party' for American White nationalism.[5] Taking place two years after the Charleston, SC church shooting,[6] the neo-Nazi-organized rally focused on the removal of confederate monuments and was a rich example of the "Nazification of the Klan," showcasing the indiscriminate merging of Nazi symbolism with Klan traditions (Ezekiel, 2002; Simi & Futrell, 2015).

White supremacist chants at the rally brought anti-Semitic rhetoric and slogans back into mainstream conversation, especially a chant thought to originate from a conspiracy theory that is also anti-Muslim.[7] Past work (Bursztyn et al., 2020; Giani & Méon, 2019; Newman et al., 2021) suggests that such inflammatory rhetoric, may expand extremist communities, leading to increases attacks on all stigmatized groups. Moreover, high-profile rallies can rapidly shift extremists' targets, potentially without even changing cumulative attack levels or emboldening or recruitment effects.

[3] For example, in January 2017, an arsonist set fire to the Victoria Islamic Center, and in September 2017 the Gates of Heaven synagogue in Wisconsin was spray-painted with swastikas and a pro-Trump message. And, in February 2017, vandals in Philadelphia toppled and desecrated at least 275 headstones at the historic Jewish Mount Carmel Cemetery, while in August 2017 the Al Maghfirah Cemetery in Minnesota was vandalized with graffiti and swastikas.

[4] https://www.splcenter.org/news/2018/10/24/splc-announces-policy-recommendations-social-media-internet-companies-fight-hate-online.

[5] https://www.vox.com/2017/8/12/16138246/charlottesville-nazi-rally-right-uva.

[6] https://www.washingtonpost.com/graphics/2020/national/confederate-monuments/.

[7] https://www.washingtonpost.com/graphics/2017/local/charlottesville-timeline/.

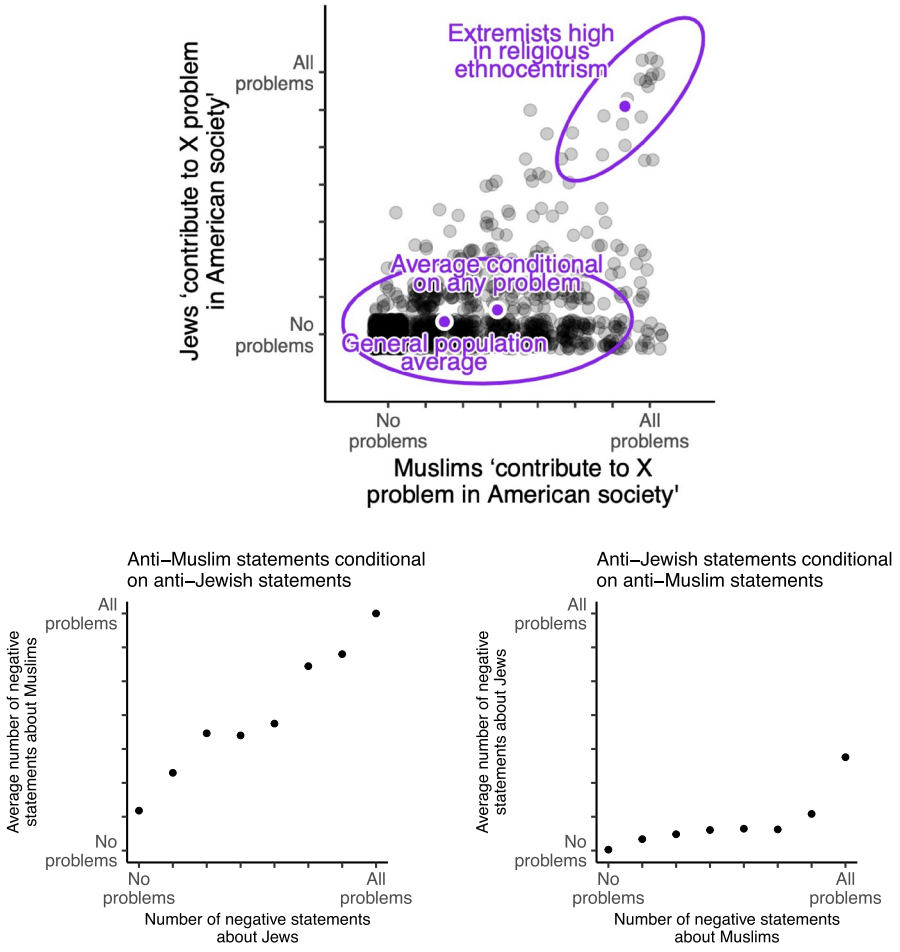

Fig. 1 Anti-Muslim and anti-Jewish attitudes among respondents in the American Mosaic Project 2014 survey. Each jittered black point in the top panel of this figure represents the number of anti-Jewish and anti-Muslim attitudes reported by a White, Non-Hispanic respondent—the sum of yes responses to seven statements, including "They don't share my morals or values" and "They want to take over our political institutions." Purple ellipses represent a normal data ellipses for responses conditional on any problem (bottom left corner) and average conditional on a sum of problems across both groups greater than 10 (top right corner). Using the same data, the bottom two panels display the average number of problems attributed to a group conditional on the number of problems attributed to the other group

In what follows, we analyze shifts in targets of online hate speech and offline hate crimes, specifically against Muslims and Jews. We test three possible manifestations of religious ethnocentrism among extremist groups:

(1) Shifts in hate targeting these two religious minority groups are positively correlated—when hate towards one group in particular declines, levels of hate towards

the other declines as well, reflecting rising and falling levels of religious ethnocentrism.

(2) Shifts in hate are uncorrelated—when hate towards one group in particular declines, we will observe no change in levels of hate towards the other, reflecting outgroup-specific hate.

(3) Shifts in hate are inversely correlated—when hate towards one group in particular declines, we will observe a change in levels of hate towards the other, reflecting constant religious ethnocentrism with target substitution.

Data

We evaluate these possibilities by constructing a large dataset on social media activity and hate crimes in the U.S. In all, our study brings together many data sources: (1) online discussions (4chan, Gab, Reddit) to study shifts in extremist rhetoric on fringe and mainstream platforms and (2) multiple hate crime data sources documenting incidents against Jews and Muslims (ADL, CAIR, FBI).[8] Turning to multiple sources allows us to verify that the patterns we observe are not specific to a single online platform or hate crime reporting system.

We begin by measuring group mentions and inflammatory rhetoric employed by extremists online. Our analyses consider mentions of racial, ethnic or religious groups on a *mainstream* social networking site (Reddit) versus *fringe* platforms (Gab and 4chan). We source data from Reddit and Gab from pushshift.io and posts from 4chan's notorious 'Politically Incorrect' subforum ('/pol/') from 4plebs.org using the Internet Archive. More information on Reddit, Gab, and 4chan can be found in SI Sections A.2.1, A.2.2, and A.2.3. Data for Gab was available for August 2016 (the beginning of the site) through August 2018. Useful for our purposes here, the structure of the site allows us to evaluate within-user shifts in hate speech targets. Data for 4chan /pol/ was available from December 2013-December 2019. Posts on 4chan are anonymous, and do not require an account to post.

To identify group mentions, we use the following keywords on each of the platforms: 'muslim', 'islam', and/or 'arab' and 'jew' and/or 'judaism'. As a comparison for the keyword approach, we also measure group mentions using a random sample of Gab and 4chan posts that we tasked workers on the crowd-sourcing website Amazon Mechanical Turk to hand label.[9] Because we do not expect extremists—or even mainstream social media users—to distinguish between Muslims and Arabs, we asked coders to label 'Muslim and/or Arab' mentions. In robustness checks (see, for example, Figures B.5 and B.8, and Table B.8), we find that hand labels, machine labels, and keywords result in the same findings. For simplicity in the main text, all aggregate analyses use the keywords; only the more fine-grained user-level analyses on Gab use the continuous machine predicted labels (see SI Section B.3).

[8] Replication data and code for this article have been posted to https://osf.io/j736h/.

[9] Figure B.3 in the SI details the full instructions given to workers.

For our social media analyses, we consider mentions of racial, ethnic or religious groups on *mainstream* (Reddit) versus *fringe* platforms (Gab and 4chan) as a measure of the salience of that group, as well as possible hate speech (on both Gab and 4chan). We focus on salience given the difficulty of labeling any specific post as hate speech without considering its broader context, and because some users may falsely claim to be a member of a group when making negative statements about it. To assess whether the vast majority of posts were negative mentions about groups, we coded the crowd-sourced hand labeled group mentions for whether the text contained an "unambiguously negative" statement about the group.[10]

Our analyses also examine whether patterns in shifts targets of hate targets in online networking sites are similar with respect to offline hate crimes. We rely on three unique databases to measure hate crimes: the FBI's Uniform Crime Reports (UCR), as well as databases assembled by the Anti-Defamation League (ADL) and the Council on American-Islamic Relations (CAIR).[11] The ADL and CAIR record anti-Jewish and anti-Muslim hate crimes respectively. Both ADL and CAIR emphasize a victim's report that bias existed in an incident to code whether the event constituted a hate crime or bias incident. And, both groups actively solicit the public to report such events to their organizations, which is a primary way in which new observations are added to the datasets. ADL also examines police reports for evidence of these events.[12]

In studying these two sources, we limit our analyses to types of bias incidents and hate crimes appearing in both data sets.[13] Our analyses begin in January 2016 because at the time of this analysis, ADL had not released complete hate and bias incident data for years prior to 2016. We then turn to the FBI data to examine whether the patterns on Gab and ADL replicate, especially for likely-to-be-reported and extreme crimes. The FBI data contains more detailed information on the types of crimes committed than ADL and CAIR. Both ADL and CAIR record physical violence, but, for example, do not distinguish types of assault, such as aggravated assaults that cause serious bodily harm or involve the use of a deadly weapon. To overcome reporting concerns,[14] we subset the FBI data to likely-to-be reported and extreme crimes when studying levels of hate crimes over time: aggravated assault, manslaughter, murder, arson, and kidnapping, since we expect these crimes to be recorded whether or not the victim trusted the police and local administration, and

[10] 90% of the posts about Jews, Muslims, and/or Arabs were considered unambiguously negative by at least one of the two coders, and 62% of the posts by both coders.

[11] We use sources beyond the FBI data because the FBI data depend on voluntary police reporting, yielding both under- and over-reporting concerns (Freilich & Chermak, 2013) In SI Figure B.11 and Table B.11, we compare these sources, and demonstrate abrupt drops in reporting to the FBI compared to both advocacy sources after the 2016 presidential election.

[12] We obtained the CAIR dataset directly from the organization and use publicly available ADL data. See https://www.adl.org/education-and-resources/resource-knowledge-base/adl-heat-map. We downloaded the FBI UCR data from their public website. See https://crime-data-explorer.fr.cloud.gov/downloads-and-docs.

[13] See SI Tables A.5 and A.6 for lists of bias incidents and hate crimes recorded by both organizations.

[14] We document sharp discrepancies between advocacy organization and government data after the 2016 election in SI Section "Comparison of ADL, CAIR, UCR Data."

we compare those crimes to the ADL and CAIR bias incidents reports. For studying week-to-week shifts in hate crimes (in differenced time series), we should not expect analyses to be so drastically influenced by major but one-off shifts in reporting and so we do not subset these analyses only to likely-to-be-reported and extreme crimes. We are also unlikely to have sufficient data on the more extreme crimes to study average rates of these crimes on a week-to-week basis.

Methods

Our analyses describe shifts in hate speech and hate crimes, as well as associations between hate speech and hate crimes. The goal of these analyses is to assess to what extent shifts in anti-Muslim and anti-Jewish hate in extremist communities might reflect shifts in religious ethnocentrism overall, versus shifts in the targets of already heightened ethnocentrism.

Our results are based on the fractions of posts that mentioned 'Muslims or Arabs' or 'Jews' by social media site. To illustrate overall versus targeted shifts in hate speech, we show (1) averages by group, (2) combined averages ('Muslims or Arabs' + 'Jews'), and (3) differenced averages ('Muslims or Arabs' − 'Jews').[15] For hate crimes, we use counts of bias incidents and hate crimes, and present the same data. Connecting these measures to the theory of ethnocentrism and target substitution, combined averages proxy for levels of religious ethnocentrism on fringe sites, while differenced averages evaluate shifts in the targets of persistently heightened religious ethnocentrism.

Our primary statistical tests use quasi-Poisson regressions (for dependent variables that are counts) and linear regressions (for fractions) on the aggregated monthly counts of group mentions or hate crimes. For Gab, where we have user-level (rather than post-only) information, we present analyses modeled without monthly aggregation and aggregate counts to the user level instead. The SI presents analyses considering within-user shifts in targeting of religious minorities, and assessing to what extent the findings presented in the main paper might change when also controlling for media coverage and terror attacks. The within-user analyses use an errors-in-variables panel regression. These findings are informative for particularly interested readers, but they do not alter the interpretation of the findings in the main text.

In comparing associations between online and offline hate, we use Granger tests (i.e., linear regressions on time series with a particular lag specification) on weekly hate speech and hate crime data. This test assesses to what extent shifts in online mentions of religious minorities are associated with shifts in hate crimes in future weeks. Using hate crime data from future weeks allows us to evaluate associations between online and offline hate that are not driven by discussion of hate crimes in the news. Although Granger tests are sometimes called Granger *causality* or sometimes *non-causality* tests, our goal here is *not* to assess whether online hate speech *causes* offline hate crimes. Instead, this merely assesses *associations* with hate crimes for two forms of expressed hate online—targeting of out-groups in general compared to shifts in the targets of ethnocentrism relative to one another.

[15] We present analyses without ratios/logging because the findings do not change when only using subtraction and addition—and prior readers have struggled to interpret the ratios and/or logged estimates.

The hate speech-hate crime association analyses use Toda–Yamamoto Granger tests, which for difference stationary time series uses a 'surplus' lag in place of explicit differencing. In this test, all weekly count variables are $\log(x + 1)$ transformed. The offline hate crimes and online hate speech models control for two weeks of prior hate crimes, terror attacks, and news coverage of hate crimes, and test for associations with two weeks of prior hate speech. We include additional information on specification selection in SI Section B.5.

Results

Inflammatory Rhetoric: Extremist Social Media Communities

Our first set of analyses examine whether a correlation exists between anti-Muslim and anti-Jewish inflammatory rhetoric across social networking sites. The time frame for our primary analyses is July–August 2017. For these social media analyses, we consider the extent to which fringe extremist communities maintained or altered mentions of religious minorities, especially Muslims and Jews. That is, once the salience of Muslims declined (as discussed in the next paragraph and demonstrated in the SI), was hate speech decline against the same group, did it decline against both groups, or did it shift toward another group (target substitution from sustained religious ethnocentrism)?

As context for these analyses, in SI Section B.1, we present findings on group salience of Muslims and White supremacy using Google Trends and news articles through the LexisNexis API. These results demonstrate that the mentions of Muslims declined in the news in July 2017, the month just prior to the Unite the Right rally,[16] perhaps leaving a space later filled by other forms of inflammatory rhetoric, such as anti-Semitic hate. We also show in SI Section B.1 that the salience of White supremacists increased the month of the Unite the Right rally, as expected from coverage of anti-Semitic chants at the rally. Thus, our analyses that follow examine whether a corresponding decline in anti-Muslim hate occurs in July 2017, and whether there is a shift in anti-Jewish hate in August 2017.

To test group salience on mainstream versus extremist online platforms, we consider several social networking websites ordered by their extremity, as measured by the proportion of posts including racial or ethnic slurs and, given that we can collect them, that are not quickly removed by moderators.[17,18] Figure 2 displays these results.

[16] Corresponding with a shift in rhetoric by Donald Trump on Twitter, see Figure B.10 in SI Section B.8, around a month after an escalation of coalition airstrikes on ISIS in Syria.

[17] See Table A.3 in the SI.

[18] Of the websites, 4chan /pol/ is the most extreme (use of a Black slur is more frequent than the word 'Black', for example), slurs appear in less than 0.1% of posts on the mainstream site Reddit (that have not been deleted by moderators or moderation bots prior to archiving), and Gab, a website advertised for "free speech" lies between the two. In these analyses, the main text figures show the ratio of 'Muslims or Arabs' to 'Jews' mentions using both keywords and supervised predictions from crowd-sourced hand labels. 'Muslim' and 'Jew' here are coded using keywords only ('muslim', 'jew', 'islam', 'judaism', ' arab ') See SI Sections B.2 and B.3 for labeling and model training details.

First, on Reddit, the *mainstream* social site we analyze, the red dotted lines in the two bottom panels in Fig. 2 display a decline in mentions of Muslims or Arabs (IRR 0.59, 95% CI 0.45–0.76), and no change in mentions of Jews (IRR 1.06, 95% CI 0.89–1.25) during the time period examined and in subreddits dedicated to politics and religion (see bottom panels).[19] As such, at least on the *mainstream* social networking site we analyze, we primarily observe shifts in discussion of Muslims, and little change in discussions of Jews.

On the fringe site Gab, however, we observe much more pronounced shifts in targets of hate against both groups. The light green solid lines in the bottom panels in Fig. 2 demonstrate that mentions of Muslims or Arabs abruptly declined to 70% of their previous level on Gab after Unite the Right (95% CI 0.64–0.77), while mentions of Jews increased to 225% of their prior level after Unite the Right (95% CI 1.99–2.54). Similar patterns persist on 4chan. The darker green solid lines in the two bottom panels demonstrate a sharp decline in mentions of Muslims or Arabs (IRR 95% CI 0.47–0.72) and a more gradual but large increase in mentions of Jews (IRR 95% CI 1.53–1.73).[20] Thus, the patterns on the two *fringe* social networking sites suggest a different pattern: inflammatory rhetoric against Jews rose in these venues as mentions of Muslims declined, suggesting a substitution effect.

To connect these findings to the target substitution hypothesis, the top panels of Fig. 2 assess whether these group-by-group shifts in mentions represent notable shifts in overall (combined) mentions of either group, or if the relative mentions of each group see larger shifts over time. In the top-left panel, we first evaluate whether the proportion of discussion about Jews and Muslims *combined* shifted on Gab or 4chan during the period studied. If they had shifted, it would indicate that the fraction of attention given to religious ethnocentrism had fundamentally altered over time, *and* might also suggest that mentions of the two groups were unrelated (e.g., we would observe a decline in this line if hate targeting one group had declined, even without any change in the other). We find that the total mentions of either 'Muslims or Arabs' or 'Jews' were high but relatively constant from early 2016 through the end of 2018, with the exception of a decline just prior to the 2016 election (during which time increased content *about the election* may have decreased the *fraction* of content that was about minority groups instead). This suggests a relatively constant level of religious ethnocentrism on fringe platforms, providing evidence for theories advanced by Kam and Kinder (2012); Kalkan et al. (2009).

Finally, and in the same vein, we also plot the *difference* between mentions of Muslims and Jews. If we observe no difference, this would suggest that the inflammatory rhetoric targeting the two declined concurrently. A negative (or positive) shift or slope, however, indicates a difference in attention being paid to one group over the other across the fringe social networking sites. And, as can be seen in the top right panel of Fig. 2, we see a noticeable shift in targets of religious ethnocentrism from Muslims or Arabs to Jews on both Gab and 4chan. These findings suggest that high levels of religious ethnocentrism may have manifested through new,

[19] See SI Figure B.4 for analyses of all 'subreddits' which demonstrate the same finding.

[20] See B.6 for a full summary table of logged coefficients.

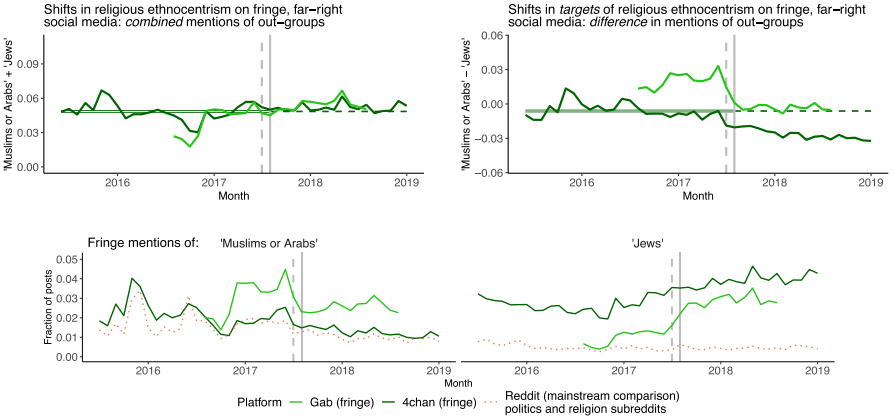

Fig. 2 The top-left panel of this figure displays the total fraction of mentions of 'Muslims or Arabs' and 'Jews' over time on Gab, 4chan /pol/ and Reddit and the top-right panel displays the difference in the fraction of specific out-group mentions. The bottom panels display fractions of mentions of 'Muslims or Arabs' and 'Jews' separately and by social media side. The horizontal green lines in the top panel indicate average values for 4chan prior to Unite the Right. On the fringe sites, overall mentions of either out-group are, by and large, stable over time despite shifts in the mentions of each group individually (Color figure online)

or reverted, targeting of Jews on fringe online social networking sites, and that this targeting replaced negative rhetoric about Muslims.

Next, we study Gab users, for whom we have individual-level data to unpack these findings, and test whether an increase in anti-Jewish speech is simply due to more users arriving to alt-right networking sites, or to increases in posts about Jews not classified as hate speech. Figure 3 displays changes in the numbers of mentions of Muslims/Arabs and Jews that also contained hate speech (compared to mention rates in January 2017) for Gab users that posted every month from January 2017 until the end of the data coverage in August 2018,[21] This establishes that the shift from anti-Muslim/Arab to anti-Jewish speech on Gab was not solely driven by an influx of new users, and that it occurs for the subset of posts labeled hate speech.

The results so far demonstrate that across two fringe social media platforms, a shift in targets of online hate from Muslims to Jews occurred during the time period studied, indicating support for the theory that hate was endemic to extremist communities and that hate towards Muslims and Jews was rooted in generalized and persistent religious ethnocentrism among White nationalists on these platforms. That we observe similar shifts on both minimally moderated sites (4chan and Gab) increases our confidence that the shift from mentions of Muslims to mentions of Jews likely generalizes across social media platforms that harbor extremists. In particular, the

[21] Hate speech labels shown in this figure were assigned using supervised models trained on data from the Gab Hate Corpus (Kennedy et al., 2018) and group mentions similarly use predicted probabilities from supervised models trained on hand labels. See SI Sections B.4, B.2, and B.3 for details.

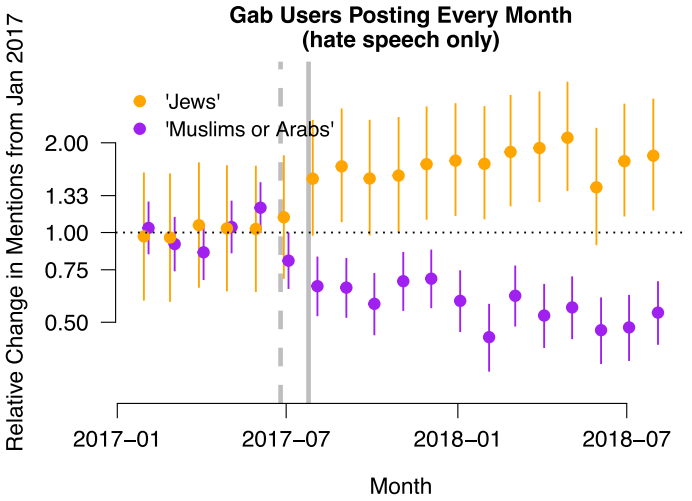

Fig. 3 This figure displays change in mentions of Jews and Muslims or Arabs that contained predicted hate speech compared to January 2017 among Gab users who posted every month January 2017 through August 2018

shift should not be attributed to the vagaries of a single platform. At the same time, the decline in anti-Muslim or anti-Arab rhetoric appears to have been particularly abrupt on both extreme and more moderate platforms.

One lingering question is whether shifts in hate happened among the same individual users. We explore this question in SI Table B.8, which displays an errors-in-variables panel regression for users on Gab (4chan is anonymous and we cannot study within-user shifts on that platform).[22] There, we find that at the individual level, month-to-month shifts in mentions of one group were positively associated with shifts in mentions of another, suggesting that users expressing religious ethnocentrism tended to target both Muslims and Jews over the course of a month. This suggests that users of fringe online social networking sites tend to target both Jews and Muslims, and provides some confirmation that these users would inhabit the top-right corner of Fig. 1 in which survey respondents possess high levels of both anti-Jewish and anti-Muslim attitudes.

However, this pattern was reversed only during the period surrounding Unite the Right. During this time, the association between mentions of 'Muslims or Arabs' and 'Jews' was negative. That is, surrounding Unite the Right, users who were less likely to mention 'Muslims or Arabs' than before were more likely to mention 'Jews' than before. This abrupt reversal at the user-level mirrors the long-term, aggregate shift in the mentions of the two groups. Even though their month-to-month shifts in mentions of the groups were still strongly correlated, both before and after the rally,

[22] Note that we do not only use a fixed effects model here because we need to evaluate associations *between* two within-user shifts of activity, and which will be measured with significant error.

their mentions of the groups tended to be *centered* on new post-"Unite the Right" levels in absolute terms.

So far, we observe a relationship between anti-Muslim and anti-Jewish hate on fringe online social networking sites, suggesting that far-right users substituted anti-Jewish animus for anti-Muslim animus in the wake of the "Unite the Right" rally. These findings not only generalize across two sites at the macro level, but also appear to persist at the individual level. Although users might *possess* both anti-Jewish and anti-Muslim attitudes, they appear to *express* these attitudes at higher or lower levels over time.

Persistence of Hate: Declines and Target Substitution in Hate Crimes

Having established that a sizable decline in mentions of Muslims across all platforms was followed by a shift to mentions of Jews on extremist sites, we next examine whether these changes in targets of online hate speech from Muslims to Jews mirror *offline* changes in hate crimes and bias incidents. Extremist leaders on fringe social media site regularly use rhetoric surrounding themes of invasion, threat, and otherness in an effort not only to increase polarization online, but also with the intention that such polarization spills into the offline space (Williams et al., 2020). Moreover, online hate speech not only intensifies and affirms feelings of in-group cohesion and outgroup hate, it also enables users to coordinate offline collective action, potentially making violence more likely (Lupu et al., 2023).[23]

We evaluate the extent to which we see combined shifts in *offline* hate crimes and bias incidents against both Muslims and Jews, and whether we see notable shifts in targeting of Jews rather than Muslims. As previously mentioned, here we use two types of data: (1) reports collected by prominent Jewish and Muslim advocacy organizations, and (2) reports of hate crimes sent to the FBI, especially violent and 'likely to be reported' hate crimes.

To begin, Fig. 4 presents findings from the two advocacy groups. The bottom panel displays the number of hate crimes and bias incidents reported to CAIR and ADL between 2016 and 2019. The figure demonstrates that violence, assault, harassment, and destruction of property directed toward Muslims declined in August 2017, indicated by the second solid grey line, (quasi-Poisson IRR 0.49, 95% CI 0.38–0.62), while aggression directed toward Jews stayed the same or increased (IRR 1.18, 95% CI: 0.88–1.57), with a possible spike in attacks in August 2017 compared to July 2017. These anti-Jewish bias incidents were higher than levels prior to the 2016 election, and largely lower than incidents just after the election and during the presidential transition. Together, then, the shifts in *offline* hate directed towards Muslims and Jews resemble those observed on fringe social media sites: anti-Jewish hate crimes appear to have increased in the aftermath of the "Unite the Right" rally while anti-Muslim hate crimes appear to have dramatically lessened in number.

[23] Not all studies posit that online hate speech will increase hate crimes (Chan et al., 2013; Glaser et al., 2002).

Next, we turn to the the top panels of Fig. 4, which mirror the top panels of Fig. 2. Here, we study both shifts in the combined hate crimes against Jews and Muslims and the difference in hate crimes. In the top-left panel, we first observe a large increase in combined number of reported hate crimes after the 2016 election, and, referencing the bottom panel, that these attacks disproportionately targeted Jews by January 2017. Next, the combined number of these hate crimes and bias incidents had begun to decline by April and May 2017. The top right-hand panel of Fig. 4 shows the difference between anti-Semitic hate crimes and anti-Muslim hate crimes, and shows a relative increase in anti-Semitic hate crimes in August 2017 (OLS log-linear change in ratio, CAIR to ADL: 0.40, 95% CI 0.31–0.52).

To substantiate the offline hate crime results, we turn to data from the two advocacy organizations and the FBI. These analyses evaluate whether these patterns of shifts in offline hate crimes replicate to another data source, where we are less likely to discover reporting artifacts—violent and 'likely to be reported' hate crimes.

Figure 5 displays the frequency of particularly violent hate crimes reported to the FBI that are either anti-Muslim/anti-Arab or anti-Jewish in nature. Specifically, we evaluate reports to the FBI of aggravated assault, murder, manslaughter, arson, and kidnapping. Here, unlike in the advocacy organization data, anti-Semitic violent crimes dramatically increase in absolute terms after the drop in anti-Muslim hate crimes in mid-2017 (IRR 1.82, 95% CI 0.99–3.42). Anti-Muslim hate crimes declined in mid-2017 (IRR 0.63, 95% CI 0.43–0.91). Notably, the effect in violent and likely to be reported crimes against Jews is much larger than the estimated effect for all bias incidents reported to ADL, though the FBI effect is estimated less precisely than the ADL effect. Full model tables for these results are shown in Table B.10.[24] In the SI, we also consider to what extent levels of reporting for less violent hate crimes shifted relative to reports to advocacy groups (see Figure B.11 and Table B.11). In all, the FBI results confirm the trends we observed in Fig. 4, and point to a clear increase in anti-Jewish hate in the wake of the "Unite the Right" rally, following a decline in anti-Muslim attacks around that period of time as well.

Lastly, we consider whether the trends in target substitution from anti-Muslim to anti-Jewish hate we observe online and offline are limited to one time point, "Unite the Right," or if they extend more broadly. Because our analyses so far study only a single event, we now test whether *week-to-week* mentions of Muslims or Arabs and Jews were associated with week-to-week shifts in offline hate crimes over several years. Broadly, we examine whether over many more time periods covering a long time horizon, the relative salience of each group appears to reflect the consequences of extremism offline better than the combined salience of religious out-groups on fringe sites.

Our analyses here specifically evaluate whether week-to-week changes in mentions of these groups on the fringe social media sites Gab and 4chan are associated with offline hate crimes and bias incidents. In this, we test associations between online mentions of out-groups with *future* hate crimes and bias incidents. This

[24] Note that this SI table displays the untransformed logged ratio coefficients from the quasi-Poisson models, rather than the exponentiated coefficients—ratios—reported here.

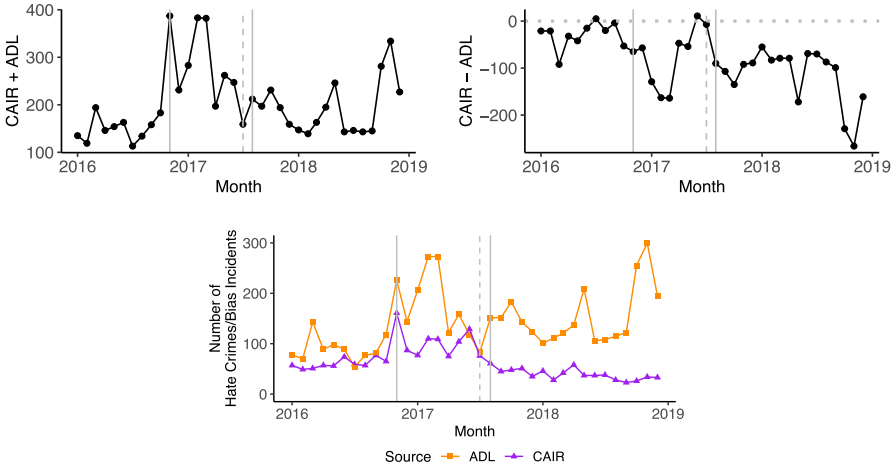

Fig. 4 The bottom panel above displays hate crimes and bias incidents recorded by ADL and CAIR by month. The top panels compares these two sources. Overall, hate crimes and bias incidents gradually decline after the 2016 election—until a very large spike in anti-Jewish incidents in late 2018. In addition to the overall patterns potentially related to the election and inauguration of Donald Trump, we also see a longer term shift in anti-Jewish hate crimes relative to anti-Muslim hate crimes in mid-2017. This second shift mirrors the activity on fringe social media sites around Unite the Right

prevents these associations from being influenced by online discussion of hate crimes covered in the news, which would not necessarily reflect targeting of these groups, especially if we were to study non-fringe platforms. We also would expect increases in expressions of hate to precede and perhaps predict violence, even when increased expressions of hate online do not themselves cause violence.

Table 1 reports the findings from the Toda–Yamamoto Granger test described in "Methods" section, using the ADL/CAIR data and FBI data, respectively.[25] This table shows that the log ratio of mentions[26] of Jews versus Muslims/Arabs is significantly associated with hate crimes and bias incidents against Jews versus Muslims after Unite the Right (and, perhaps, not before), including in models that control for the number of terror attacks with 'Muslim' or 'Islam' in attack descriptions from the Global Terrorism Database[27] and the ratio of news articles (as in other analyses, mentions of Jews to Muslims) about hate crimes.

Across both platforms, which cover different time horizons, we find no such correlation for combined mentions of these groups. In other words, ratios of mentions on fringe sites appear to reflect the incidence of hate crimes (and, given the lagged

[25] These models use the full FBI data because included lags prevent the abrupt shifts in hate crime reporting in late 2016 and early 2017, as documented in SI Section B.10.3 and which suggests potential long-term under-reporting of simple assaults and vandalism after the 2016 election and 2017 inauguration, from meaningfully influencing the models.

[26] As noted in the social media analysis section, we do not log variables in visualizations only to increase their accessibility to readers. The log ratio of mentions is a transformation of these count variables.

[27] https://www.start.umd.edu/gtd/.

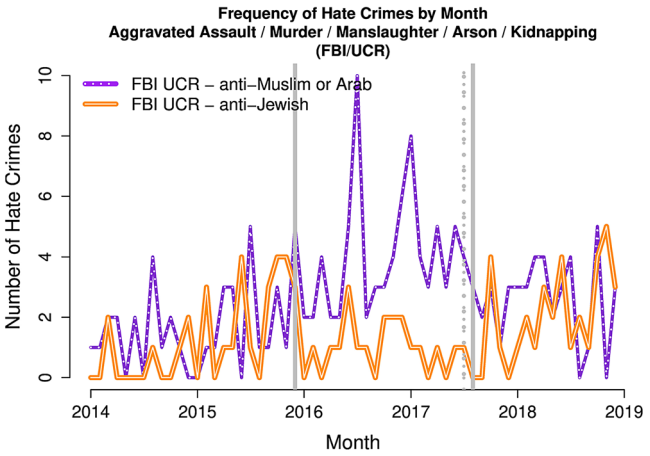

Fig. 5 Anti-Muslim or Arab and anti-Jewish aggravated assault, murders, manslaughter, arson, and kidnapping recorded in the FBI UCR data by month

online social media variables, not merely the discussion of them), while combined mentions on these sites do not.[28] Note that this does not evaluate the importance of high and constant levels of religious ethnocentrism overall—instead, these tests only evaluate to what extent changes in online activity are associated with *changes* in hate crimes over this period, and we also did not find noticeable variation in changes in combined levels of mentions of Muslims and Jews on fringe sites online. These findings provide further support for a robust relationship between anti-Muslim and anti-Jewish hate, extending beyond the Charlottesville rally.

Discussion

Since the 2016 presidential election, the U.S. has witnessed an uptick in far-right extremism and violent hate crimes. This era saw a rise in White supremacist events, such as rallies and demonstrations, growth in extremist social media forums, and a near tripling in White supremacist propaganda from 2017 to 2018.[29] But despite more research on online and offline extremism (e.g. Cikara et al., 2022; Ivandicet al., 2022; Siegel et al., 2021), much remains unknown about its over-time dynamics and consequences, particularly for religious minorities.

In this study, we find that while White nationalism was on the rise, anti-Muslim hate speech declined one month before the Unite the Right rally. However, hate remained endemic within online fringe communities, with anti-Muslim hate replaced by anti-Jewish hate speech. Additional analyses suggest that shift in the target of hate, from Muslims to Jews, was even evident among the same users. Furthermore, we find a contemporaneous and corresponding shift in offline hate crimes as

[28] Note the p-values in these models do compare the difference in mentions with the combined mentions.
[29] https://www.adl.org/resources/reports/white-supremacists-step-up-off-campus-propaganda-efforts-in-2018.

Table 1 These two tables compare associations with reported hate crimes for overall versus targeted mentions of 'Muslimsor Arabs' and 'Jews' on Gab and 4chan from August 2016–December 2018, and, in the FBI data, for 4chan fromDecember 2015 through December 2018 (We use 4chan here because it covers a longer period than the Gabdata.).

Assessing shifts in online religious ethnocentrism and shifts in targets of religious ethnocentrism: associations with bias incidents and hate crimes

Bias incidents and hate crimes in next week		ADL-CAIR		ADL-CAIR
Shifts		Combined		Difference
Combined fringe mentions		$p = 0.94$		
Difference in fringe mentions				$p = 0.03$
Hate crimes in next week	FBI	FBI	FBI, before Unite the Right	FBI, after Unite the Right
Shifts	Combined	Difference	Combined	Difference
Combined 4chan mentions	$p = 0.70$		$p = 0.44$	
Diff. in 4chan mentions		$p = 0.54$		$p = 0.01$

Each model controls for media coverage of hate crimes and terror attacks; however, these controls did not meaningfully alter the associations. Note that the Gab/4chan—ADL/CAIR (top right) and the *after* 'Unite the Right' 4chan—FBI data analyses (bottom far right) rely on similar timeframes and likely similar data

well. These corresponding shifts were found for both levels of anti-Jewish and anti-Muslim hate crimes before and after Unite the Right, as well as shifts on a *week-to-week* basis in 2017 through 2018.

Our work triangulates results from multiple social networking sites and hate crime databases, illustrating that hate does not evaporate with the decline of group salience or inflammatory rhetoric. Instead, it manifests differently within extremist communities, where it appears to persist without extensive mainstream propagation. That hate crimes could decline and shift in this way suggests that perpetrators of hate can be traced to the same communities and that they shift their targets and pretexts for hate over time.

Our study broadens our understanding of how Muslims and Jews have become interchangeable subjects of hatred for U.S. white supremacy groups. In supporting analyses, we build on previous work by Gerteis and Rotem (2022), and reveal a unique subset of Americans high in anti-Muslim and anti-Jewish animus. Our results suggest that hate is *inversely* linked over time, despite strong positive associated in cross-sectional surveys. In other words, when hate toward one group declines (Muslims), we observe a substitution in toward another (Jews).

Note that our study does not explain the universe of potential drivers of hate crimes. Past work has found that inflammatory rhetoric on mainstream news and social media platforms is associated with hate crimes overall, and our analyses illustrate some patterns of offline hate crimes that should not be attributed to differential shifts in hate against Muslims and Jews. But overall shifts too are not the sole dynamic behind hate crimes, and neither should we consider hate crimes

against different groups solely on a group-by-group basis. For fringe extremists specifically, and communities of 'endemic' hate, we demonstrate separable and distinct patterns of hate speech, and show that these patterns also reflect associations with offline hate crimes.

Though our research provides valuable insights, several questions remain unanswered. Is activity in extremist networking groups is a harbinger of offline hate, or do these online communities directly incite offline hate too? Moreover, how unique are the circumstances that might make substitution effects more likely, such as the Unite the Right rally, migration to fringe social media sites, and the early Trump presidency? Given our focus on extremists on fringe social media sites and a specific time period in the United States, we are necessarily limited in our claims about the extent of target substitution in other contexts. And, once substitution effects take hold, how much longer can they persist in fringe settings without mainstream reinforcement? Future research can further explore these questions, potentially focusing on other hate group targets and meeting events.

From this research, we learn about the tenacity of hate and the shifting nature of its targets. For instance, focusing only on decreased hate toward Muslims would provide an incomplete narrative, and the story would have been a positive one: online and offline hate towards Muslims decreased through 2017. But, expanding the research lens to another religious outgroup that similarly threatens white supremacists allows us to understand that the situation has worsened for American Jews. Thus, a seemingly increasing tolerance towards one group does not necessarily reflect a broader trend. Importantly, our findings suggest that expressions of hate can be traced to the same extremists.

We hope this research informs future studies, particularly those seeking to understand persistent and *shifting* hate among the same communities and individuals. Our work offers an initial exploration of long-term target substitution effects. But future work could broaden this understanding by examining different time points, targets, and contexts. Given the significant correlations found here between online extremist discussions and offline intergroup conflict, it is crucial for scholars to further examine these relationships.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11109-023-09892-9.

Data Availability Interview guides are available as supplementary material.

Declarations

Conflict of interest Replication materials for this article have been posted to: https://osf.io/j736h/.

References

Allport, G. (1954). *The nature of prejudice*. Addison-Wesley.

Baker, J., Perry, S., & Whitehead, A. (2020). Keep America Christian (and White). *Sociology of Religion, 81*(3), 272–293.

Blalock, H. (1967). *Toward a theory of minority-group relations*. Wiley.

Bloom, M., Tiflati, H., & Horgan, J. (2019). Navigating ISIS's preferred platform: Telegram. *Terrorism and Political Violence, 31*(6), 1242–1254.
Boydstun, A. (2013). *Making the News*. University of Chicago Press.
Bursztyn, L., Egorov, G., & Fiorin, S. (2020). From extreme to mainstream. *American Economic Review, 110*(11), 3522–3548.
Chan, J., Ghose, A., & Seamans, R. (2013). The Internet and Hate Crime. Technical Report.
Cikara, M., Fouka, V., & Tabellini, M. (2022). Hate crime towards minoritized groups increases as they increase in sized-based rank". *Nature Human Behaviour, 6*, 1537–1544.
Dunwoody, P., & McFarland, S. (2018). Support for anti-Muslim policies. *Political Psychology, 39*(1), 89–106.
Erikson, K. (1966). *Wayward Puritans*. Wiley.
Ezekiel, R. (2002). An ethnographer looks at neo-Nazi and Klan groups. *American Behavioral Scientist, 46*(1), 51–71.
Fouka, V., & Tabellini, M. (2022). Changing in-group boundaries. *American Political Science Review, 116*(3), 968–984.
Freilich, J., & Chermak, S. (2013). *Hate crimes*. US Department of Justice, Office of Community Oriented Policing Services.
Gerteis, J., & Rotem, N. (2022). White anti-Semitic and anti-Muslim views in America". *The Sociological Quarterly, 64*(1), 144–164. https://doi.org/10.1080/00380253.2022.2045882
Giani, M., & Méon, P.-G. (2019). Global racist contagion following Donald Trump's election. *British Journal of Political Science, 1*, 1–8.
Glaser, J., Dixit, J., & Green, D. (2002). Studying hate crime with the Internet. *Journal of Social Issues, 58*(1), 177–193.
Hafez, M., & Mullins, C. (2015). The radicalization puzzle. *Studies in Conflict & Terrorism, 38*(11), 958–975.
Hawdon, J., Bernatzky, C., & Costello, M. (2019). Cyber-routines, political attitudes, and exposure to violence-advocating online extremism. *Social Forces, 98*(1), 329–354.
Hobbs, W., & Lajevardi, N. (2019). Effects of divisive political campaigns on the day-to-day segregation of Arab and Muslim Americans. *APSR, 113*(1), 270–276.
Isom, D., Mikell, T., & Boehme, H. (2021). White America, threat to the status quo, and affiliation with the alt-right. *Sociological Spectrum, 41*(3), 213–228.
Ivandic, R., Kirchmaier, T., & Machin, S. (2022). International terror attacks and local out-group hate crime. Working Paper.
Jardina, A., & Stephens-Dougan, L. (2021). The electoral consequences of anti-Muslim prejudice. *Electoral Studies, 72*, 102364.
Kalkan, K., Layman, G., & Uslaner, E. (2009). "Bands of others''? Attitudes toward Muslims in contemporary American society. *The Journal of Politics, 71*(3), 847–862.
Kam, C., & Kinder, D. (2012). Ethnocentrism as a short-term force in the 2008 American Presidential Election. *AJPS, 56*(2), 326–340.
Kennedy, B., Atari, M., Mostafazadeh Davani, A., Yeh, L., Omrani, A., Kim, Y., Koombs, K., Havaldar, S., Portillo-Wightman, G., Gonzalez, E., Hoover, J., Azatian, A., Hussain, A., Lara, A., Olmos, G., Omary, A., Park, C., Wang, C., Wang, X., … Dehghani, M. (2018). The Gab Hate Corpus: A collection of 27k posts annotated for hate speech. *PsyArXiv*. https://doi.org/10.31234/osf.io/hqjxn
Kyler, A., & Charron-Chénier, R. (2022). Understanding alt-right interest using Internet search data. *Social Science Research, 106*, 102729.
Lajevardi, N. (2020). *The politics of American Islamophobia*. Cambridge University Press.
Lajevardi, N., & Abrajano, M. (2019). How negative sentiment toward Muslim Americans predicts support for Trump in the 2016 Presidential Election. *The Journal of Politics, 81*(1), 296–302.
Lupu, Y., Sear, Richard, V., Nicolas, Leahy, R., Restrepo, N. J., Goldberg, B., & Johnson, N. F. (2023). Offline events and online hate. *PLoS ONE 18*(1), e0278511.
Mason, L., Wronski, J., & Kane, J. (2021). Activating animus. *American Political Science Review, 115*(4), 1508–1516.
Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of hate speech in online social media. In *Proceedings of the 10th ACM conference on web science* (pp. 173–182).
McVeigh, R. (2009). *The rise of the Ku Klux Klan* (Vol. 32). University of Minnesota Press.

Meuleman, B., Abts, K., Slootmaeckers, K., & Meeusen, C. (2019). Differentiated threat and the genesis of prejudice: Group-specific antecedents of homonegativity, Islamophobia, anti-Semitism, and anti-immigrant attitudes. *Social Problems, 66*(2), 222–244.

Müller, K., & Schwarz, C. (2023). From hashtag to hate crime: Twitter and antiminority sentiment. *Journal of the European Economic Association, 19*(4), 2131–2167.

Müller, K., & Schwarz, C. (2023). Social media and hate crime. *AmericanEconomic Journal: Applied Economics, 15*(3), 270–312.

Newman, B., Merolla, J., Shah, S., Lemi, D.C., Collingwood, L., & Ramakrishnan, K. (2021). The Trump effect: An experimental investigation of the emboldening effect of racially inflammatory elite communication. *British Journal of Political Science, 51*(3), 1138–1159.

Nithyanand, R., Schaffner, B., & Gill, P. (2017). Online political discourse in the Trump era. arXiv preprint. arXiv:1711.05303

Olzak, S. (1994). *The dynamics of ethnic competition and conflict*. Stanford University Press.

Oskooii, K., Dana, K., & Barreto, M. (2021). Beyond generalized ethnocentrism. *Politics, Groups, and Identities, 9*(3), 538–565.

Penning, J. (2009). Americans' views of Muslims and Mormons. *Politics and Religion, 2*(2), 277–302.

Saleem, M., Prot, S., Anderson, C., & Lemieux, A. (2017). Exposure to Muslims in media and support for public policies harming Muslims. *Communication research, 44*(6), 841–869.

Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., & 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), 71–104.

Simi, P., & Futrell, R. (2015). *American Swastika: Inside the white power movement's hidden spaces of hate*. Rowman & Littlefield.

Stangor, C., Sullivan, L. A., & Ford, T. E. (1991). Affective and cognitive determinants of prejudice. *Social Cognition, 9*(4), 359–380.

Stimson, B. (January 10, 2021). Gab gaining 10,000 users per hour, CEO claims, after Trump's permanent Twitter suspension. *Fox Business News*.

Tesler, M. (2022). President Obama and the emergence of Islamophobia in mass Partisan preferences. *Political Research Quarterly, 75*(2), 394–408.

Thompson, J. (2021). What it means to be a "true American''. *Nations and Nationalism, 27*(1), 279–297.

Walther, J. B. (2022). Social media and online hate. *Current Opinion in Psychology, 45*, 101298.

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology, 60*(1), 93–117.

Wlezien, C. (2023). News and public opinion. *The Journal of Politics*. Forthcoming.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Zick, A., Küpper, B., & Hövermann, A. (2008a). *Intolerance, prejudice and discrimination—A European report*. Friedrich-Ebert-Stiftung Forum Berlin.

Zick, A., Wolf, C., Küpper, B., Davidov, E., Schmidt, P., & Heitmeyer, W. (2008b). The syndrome of group-focused enmity. *Journal of Social Issues, 64*(2), 363–383.

Authors and Affiliations

William Hobbs[1] · Nazita Lajevardi[2] · Xinyi Li[1] · Caleb Lucas[3]

✉ William Hobbs
hobbs@cornell.edu

✉ Nazita Lajevardi
nazita@msu.edu

Xinyi Li
xl624@cornell.edu

Caleb Lucas
clucas@rand.org

[1] Cornell University, Ithaca, USA

[2] Michigan State University, East Lansing, USA

[3] RAND Corporation, Santa Monica, USA