Journal of Computational Social Science (2024) 7:1305–1325
https://doi.org/10.1007/s42001-024-00272-9

RESEARCH ARTICLE

Hostility on Twitter in the aftermath of terror attacks

Christian S. Czymara[1,2] · **Anastasia Gorodzeisky**[1]

Received: 25 November 2023 / Accepted: 17 March 2024 / Published online: 27 April 2024

Abstract

This study investigates the relationship between major Jihadist terror attacks and manifestations of ethno-religious hostility on social media. Analyzing approximately 4.5 million time-stamped Tweets from 1.2 million users across five European countries, the study focuses on content discussing migration and related topics in the weeks before and after ten significant terror attacks. The findings show a notable and robust increase in hostile Tweets after an attack. An interrupted time series analysis demonstrates a 10% point surge at the time of the attack, followed by a gradual decline. Accordingly, the impact of such attacks on online hostility diminishes approximately seven days after the event. Further analyses reveal that while attacks have the strongest effect on Tweets about Muslims and Islam, the attacks also increase hostility in Tweets about migration in general. We find that the overall attack effect is driven by both intra-user changes in Tweeting and changes in the composition of users posting after an attack. The findings underscore the importance of understanding the interplay between terrorist events and online discourse, shedding light on the dynamics of ethno-religious hostility in the digital realm.

Keywords Terrorism · Migration · Muslims · Hostility · Hate speech · Incivility · Attitudes · Twitter · X · Europe

✉ Christian S. Czymara
czymara@tauex.tau.ac.il

[1] Tel Aviv University, Tel Aviv-Yafo, Israel

[2] Goethe University Frankfurt, Frankfurt am Main, Germany

Introduction

Several major terror attacks conducted in the name of political Islam have hit Europe in recent years[1], some of these attacks resulting in dozens of fatalities. In addition to the devastating consequences for those directly affected, such attacks also have an impact on inter-group relations in European societies, in that public speakers and political elites sometimes cite mass immigration as a contributory factor [2, 3]. In turn, Muslims in Europe have reported a rise in perceived discrimination [4] and a decline in mental health and wellbeing [5] after Jihadist terror attacks. On the side of the majority population, existing evidence regarding the impact of terrorism on attitudes toward immigration and immigrants is mixed. While some studies found no evidence for a relationship between terrorist attacks and such attitudes [e.g.: 6,7], others have shown more negative public attitudes toward immigrants [e.g.: 8–13] or higher exposure to online hate speech [14] after attacks. Most of these studies are based on survey data and consequently have methodological limitations, such as being too coarse to measure the time-sensitive effects, being restricted to predefined survey items, or social desirability bias [see 15]. Complementing existing survey research, this paper sets out to overcome these limitations by studying changes in ethnic and religious hostility expressed on social media[2]. For the context of the study, we define ethno-religious hostility as aggressive, antagonistic, or negative interactions and rhetoric aimed at ethno-religious minorities or migrants, often characterized by prejudice, xenophobia, and discrimination. Specifically, we ask: *Does ethno-religious hostility on social media increase after terror attacks?* We use time-stamped data capturing commenting behavior in over 4.5 million Tweets, posted by about 1.2 million different users across five European countries, before and after ten major terror attacks.

In doing so, this study also speaks to the literature on the stability of attitudes. Scholars have recently argued that attitudes and values are mostly stable over the life course [17, but see 18]—including attitudes toward immigration [19]. Following this reasoning, attitudes are largely formed during the formative years of adolescence and early adulthood [20, 21]. While a life course perspective is clearly beyond the scope of the present study, we have set out to test whether sudden events of terrorism contribute to short-term changes in expressions of hostility by individual users of a social media platform, and how fast these changes abate. The advantage of this approach lies in analyzing opinions and their shifts at the precise moment a person decides to express them (i.e., chooses to Tweet) rather than relying on scheduled interviews using pre-set questionnaires, as with survey data collection.

Beyond methodological benefits in studying changes in ethno-religious hostility expressed on a social media platform, Twitter is a meaningful field of study for social scientists given its positioning as a "digital socioscope" [22 (p. 1)] or "online

[1] When we speak of terrorism or attacks in this paper, we refer to terror attacks with a Jihadist/Islamist motivation unless stated otherwise. While other types of terrorism can also be important for inter-group relations in Europe [1], they do not fit our theoretical framework as outlined below.

[2] This particularly concerns Muslims, clearly a religious group. Yet within the European context, a Muslim identity can also have an ethnic dimension, with some scholars using terms such as *Islamic ethnicity* [16]. For this reason, we also refer to our outcome as *ethno-religious* hostility.

marketplace of ideas" [23], one that plays an important role for public and political discourse. Social science research has recently come to appreciate Twitter data as a means of testing theories that hitherto had mainly been examined empirically in the context of survey data analysis. In the area of public attitudes toward immigration and immigrants, a pioneering example is the study of Menshikova and van Tubergen [24], who tested the predictions of group threat theory (see below) using Twitter. Mirroring findings from attitudinal survey research [25–27], Menshikova and van Tubergen [24] reported more negative Tweets on immigration when the topic is salient in the news. Following this path, we employed Twitter to test the impact of terrorist attacks on ethno-religious hostility. While this effect has been extensively studied in the context of survey research [for example: 8–13], it has received limited attention in research on social media. Our goal is to analyze, from a computational social science perspective, the much-debated hypothesis that terrorism shapes anti-immigrant attitudes.

More broadly, analyzing shifts in social media debates is also important because hostile discussions on social media can influence the reporting of traditional news outlets, as suggested by intermedia agenda-setting theory [28, 29]. Likewise, online language use can have consequences with respect to offline behavior. For example, the level of online hate speech has been shown to correlate with the vote share of far-right parties [30]. Even more dramatically, hate speech against refugees on Facebook has also predicted anti-refugee violence in Germany [31].

The impact of Islamist terrorism on ethno-religious hostility

Group threat theory contends that negative attitudes and hostile behavior toward out-groups (e.g., immigrants, ethnic and religious minorities) are the result of the impression that minority members pose a threat to the interests of a given society's majority group [32, 33]. In his seminal paper on racial prejudice as a sense of group position, Herbert Blumer [34] argues that important events play a key role in the development of exclusionary and discriminatory views toward out-group populations because such events contribute to a particular majority's perception of the minority group and intergroup relations. Blumer argues that events "loaded with great collective significance", which feed into representations of the minority group as a threat in public discussions, are "particularly potent in shaping the sense of social position" [34 (p. 6)]. Terror attacks certainly fall into the category of such significant events, as they pose a threat to individual and collective safety as well as to social norms, and are often blamed on immigrants in general and Muslim immigrants in particular. Thus, terror attacks in the name of political Islam can have the power to shape attitudes toward immigrants and Muslims.

The theoretical prediction of group threat theory — that terror attacks lead to more exclusionary attitudes among the majority population — has been tested extensively by social scientists. This body of research can be grouped into three categories: Natural experiments based on secondary survey data; studies based on self-collected survey data; and papers drawing from digital trace data.

The first set of studies exploits the fact that several large survey programs happened, unintentionally, to be active in the field at the time of such terror attacks. The timing of these attacks, being exogenous, resembles a natural experiment. Based on such a design, several studies have reported more unfavorable sentiments toward immigrants after attacks [5, 8–13, 35, 36]. However, the claimed magnitude of this impact varies across cases. Some studies even reported null effects [6, 7, 37, 38].

A second set of studies is based on self-collected data specifically designed to capture the way people perceive the shifts in social media debates following terror attacks. For example, Kaakinen, Oksanen, and Räsänen [14] showed that young Finns reported being exposed significantly more often to online hate speech after the November 2015 Paris attacks; Oksanen and colleagues [39] demonstrated that exposure to online hate speech after this attack correlated with the perception that one's society is characterized by fear. Based on a self-designed online forum that systematically varied existing comments, Álvarez-Benjumea and Winter [40] showed that expressions of online hate speech toward refugees after terror attacks tend to be conditional on the normative context of a platform.

Very few studies directly examined the impact of terror attacks on online behavior by drawing from digital trace data (i.e., social media data analyses). Employing topic modeling on 51,000 German Tweets posted in the week after the Berlin attack of 2016, Fischer-Preßler, Schwemmer, and Fischbach [41] found that Tweets expressed sympathy for victims but also nationalistic sentiments. Czymara et al. [42] showed, based on over 100,000 YouTube comments on immigration-related issues before and after a set of attacks, that ethnic insulting was more present after attacks. Analyzing 11 events related to terrorism and crime, Giavazzi et al. [30] showed the language used on German Twitter became more similar to the language used by the right-wing Alternative für Deutschland (AfD) party after such events. Finally, Jović et al. [43], analyzing views of Wikipedia pages before and after attacks, demonstrated that attacks boosted attention to pages on content related to the attackers (such as terrorism or Islam), as well as to pages on security and self-perception (such as the national or religious identities of one's society). This underscores the importance of Islamist attacks for processes of boundary-making in Western countries. In sum, these considerations lead us to expect that:

Hypothesis 1 Hostility on social media increases after terror attacks.

However, the effect of terrorism on social media debates is likely to change over time. Prior research proposes that online communication in times of crisis follows Pennebaker's model [44], which consists of three stages: emergency, inhibition, and adaptation [45]. The emergency phase features heightened anxiety and more talk about the crisis event, followed by a decrease in discussion during the inhibition phase, and finally, the adaptation phase signifies a return to normalcy, indicating the community has psychologically moved past the incident [45]. In the context of the Paris terrorist attacks in November 2015, Garcia and Rimé, [46] show an immediate surge in negative affect in Tweets, including anxiety, sadness, and anger, after the attack, which leveled off within several days. However, the rise in increased solidarity after the attack was long-term [46] - a pattern consistent with the response to other types of disasters [47]. Attitudinal survey studies have also shown that the potential effects of terrorism on hostility tend to be short-lived [10] and abate over time [11]. Turkoglu and Chadefaux [38] reported that the effects of attacks on attitudes toward immigration disappear within two weeks. Similarly, Helbling and Meierrieks [48] concluded in their review paper that the effects of terrorism on anti-immigration attitudes tend to be short-lived. We expect that online discussions on immigration-related topics after terrorist attacks follow a similar trajectory as in [45, 46]. This implies that the immediate increase in hostility levels should weaken over several days and, ultimately, return to baseline levels. We formulate the following hypothesis:

Hypothesis 2 The effect postulated in H1 levels off over several days.

On a conceptual level, it is likely that Jihadist terror attacks do not affect views and behavior toward all immigrants equally. Rather, prejudice is dependent on the degree to which a group is seen as threatening within a particular social context [49]. People differentiate between different minority groups when forming their opinions; as a consequence, attitudes toward different groups are associated with different predictors [50, 51]. Jihadist attacks are perceived as most directly connected to Muslims because such attacks are motivated by political Islam. Thus, views toward Muslims are most likely to be affected by such attacks [52]. However, there might also be a generalization toward public views of immigration more broadly. This is because terrorism is sometimes blamed on immigration in general (see [48] for an overview). Viktor Orbán, Hungary's right-wing prime minister, referenced this in a comment that "all the terrorists are migrants" [2]. We hence expect a specific as well as a generalized component in the increase of ethno-religious hostility after attacks.

Hypothesis 3a Islamist attacks have the strongest effect on Tweets about Muslims.

Hypothesis 3b There is a spillover effect onto Tweets about immigration in general, even if such Tweets do not mention Muslims or Islam.

Finally, we want to test the mechanisms that cause the rise of hostility on social media after attacks. The argumentation of group threat theory implies an attitudinal change in responses to threatening events such as terrorist attacks [11]. However, some research has questioned whether immigration attitudes react to such shocks [19]. More broadly, Kiley and Vaisey [17] argue that "personal culture", i.e., an individual's "attitudes, worldviews, values, dispositions, and associations" [17 (p. 478)], is mostly stable over one's life course [but see 18 for a counterargument]. If this is true, any observed differences over time should be driven by disparities between people expressing their opinions at different time points—that is, interpersonal differences that perhaps appear due to socialization processes [20, 21]—and not because people revise their views and change their behavior. In the context of this study, this means that higher levels of ethno-religious hostility after terror attacks would be driven by differences in the composition of users who post before and after an attack. Put differently, more hostile commenters join the discussion due to an attack while more peaceful ones may leave. The competing hypothesis is that individual users actually update their views— and thus become more hostile in their Tweets.

Existing research on social media data mainly favors the explanation of composition effects [42, 54, 55]. However, these studies have not provided a sensitive test of within-user changes in posting on social media platforms following a terror attack. Flores [54], for example, did not examine terrorist attacks but rather the influence of the introduction of an anti-immigration law in the US on Twitter. In contrast, Spörlein and Schlueter [55] and Czymara et al. [42] both analyzed data from YouTube comments on immigration issues made before and after terrorist attacks. However, YouTube is structurally different from Twitter: while debates on Twitter are often highly interactive, users might post only a single comment under a given video on YouTube. This complicates the identification of changes in the content of individual user's comments on YouTube. Czymara et al. [42] reflected this by observing that "most users who comment before a terror attack do not comment after an attack, which renders a general statement on within-user change complicated" [42 (p. 546)]. We believe that Twitter's vivid and interactive culture makes it an ideal platform to test the question of whether individual users actually become more hostile on social media due to significant events. Thus, to understand the sources of the rise in ethno-religious hostility in the aftermath of terrorist attacks, we ask whether individual users became more hostile, and formulated the following hypothesis:

Hypothesis 4 Within users' hostility increases after terror attacks.

To put Hypothesis 4 in other words, the Tweets of the same users will be more hostile after attacks than before attacks.

Data and method

Selection of cases

We draw upon Tweets dealing with immigration and related topics such as refugees, asylum, Muslims, and Islam, posted in the week before and after ten major Islamist terror attacks in Europe.[3] We included all major terrorist attacks that took place in Europe after the establishment of Twitter, defined as attacks that killed at least five and injured more than ten people.[4] For each attack, we removed Tweets posted in the first 30 min after the attack, to ensure that information about the attack was already circulating and to ensure that Tweets analyzed were in reaction to the event in question. Table 1 gives an overview of the attacks that were included in the analysis. All code is available on the Open Science Framework: https://osf.io/zdt5b.

[3] The January 2015 Paris attacks spanned more than one day. Because we collected data up until seven days after the end of each attack, this resulted in a somewhat longer period in this particular case.

[4] We also downloaded Tweets posted during the Vienna shootings on 02 November 2020. Closer inspection revealed, however, that a Jihadist stabbing in Nice on 29 October 2020 corrupted our design.

Table 1 Overview of attacks

City	Date	Attack	Deaths	Injuries	Abbreviation
Paris (FR)	7–9 January 2015	Charlie Hebdo+kosher supermarket shootings	17	19	FR15_CH
Paris (FR)	13 November 2015	Bataclan shootings and other attacks	137	413	FR15_BATA
Brussels (BE)	22 March 2016	Airport and metro station bombings	32	340	BE16
Nice (FR)	14 July 2016	Bastille Day truck attack	87	270	FR16
Berlin (DE)	19 December 2016	Christmas market truck attack	13	48	DE16
London (UK)	22 March 2017	Westminster Bridge attack	5	48	UK17_west
Manchester (UK)	22 May 2017	Manchester Arena bombing	22	239	UK17_man
London (UK)	3 June 2017	London Bridge attack	8	48	UK17_lon
Barcelona (ES)	17 August 2017	Van attacks	15	104	ES17
Strasbourg (FR)	11 December 2018	Christmas market attack	5	11	FR18

All Tweets were collected between December 2022 and January 2023 from the v2 API endpoint for the Academic Research Product Track using [56]. The goal was to identify original Tweets concerned with immigration-related topics. To this end, we used a language-specific search string including keywords such as “immigration”, “foreigners”, and “refugees”. Tweets were collected when they included at least one of these keywords. Moreover, we included the terms “Muslims” and “Islam” because, in the European context, Muslims are quite possibly the most debated immigrant group.[5] However, we excluded terms that have a strong connotation with terrorism (e.g., “Jihad”, “Islamist”). Likewise, we did not directly search for insults, derogatory terms, and neither for attack-specific hashtags, as these are only plausible for the period after an attack. For each term, we searched for its *direct appearance* in a Tweet as well as its *hashtag*. The English search string was translated into French, German, and Spanish by native speakers. We used language-specific queries for each case but decided against restricting our sample to accounts sharing their locations. When applicable, we used singular, plural, male, and female versions of the relevant terms. To test Hypotheses 3a and 3b, we re-ran the models excluding those Tweets that mentioned terms relating to Muslims or Islam. We did not include Retweets in our sample as they do not constitute original content. This strategy yielded a data set of 4,596,300 Tweets.

Generally, Tweets about migrants, refugees, and Muslims increased strongly after attacks but leveled off over time. Fig 1 displays this trend for the pooled data by showing the number of Tweets per minute before and after an attack. There is a clear spike shortly after the attacks, where the aggregated number rises to over 850 Tweets per minute. This trend then levels off over time. Table 2 shows that the data are skewed toward the three cases from the UK, which make up about 75% of the overall

[5] The English terms are: “migration”, “immigration”, “migrant”, “immigrant”, “foreigner”, “migrants”, “immigrants”, “foreigners”, “refugee”, “refugees”, “asylum”, “islam”, “muslim”, “muslims”. All search strings and settings are accessible at https://osf.io/zdt5b.

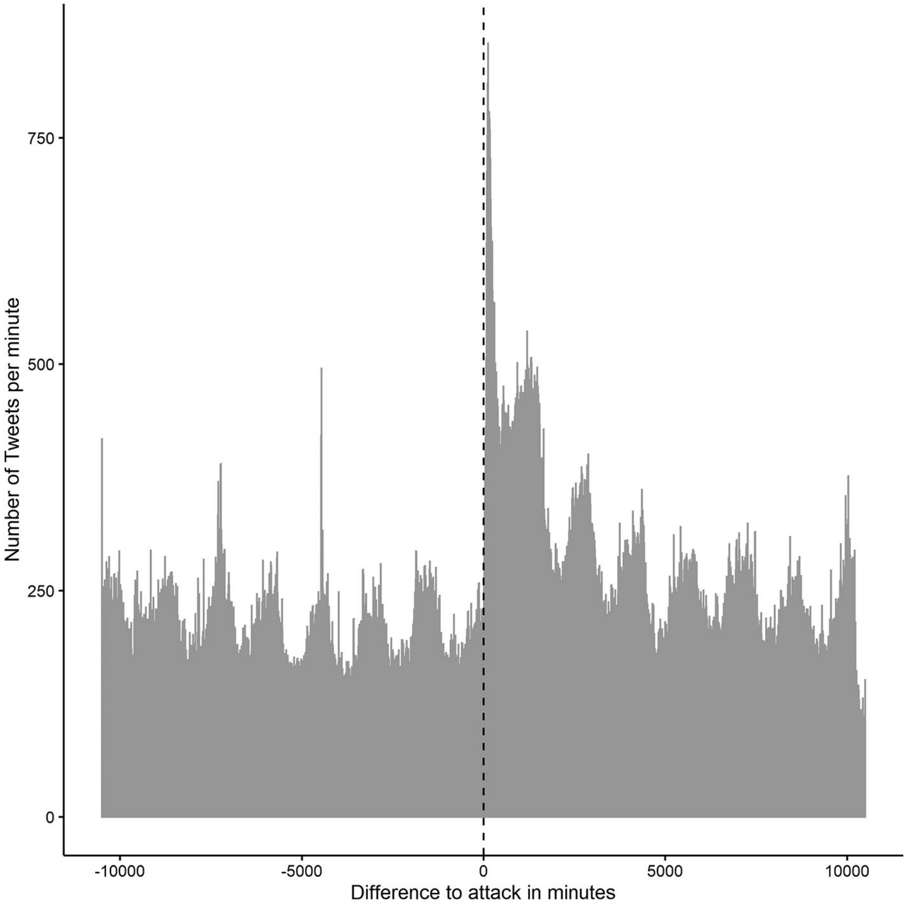

Fig. 1 Global distribution of Tweets

Table 2 Distribution of Tweets by case

	Week Before	Week After
BE16	36,515	59,578
DE16	19,799	22,628
ES17	69,233	161,862
FR15_BATA	38,436	127,831
FR15_CH	28,005	202,412
FR16	30,850	69,499
FR18	50,991	49,367
UK17_lon	560,621	650,083
UK17_man	489,011	694,636
UK17_west	619,634	615,309
Total	1,943,095	2,653,205

data. This is to be expected though; the English-speaking public is by far the largest one on Twitter. The fewest Tweets came from the German attack, which makes up 0.9% of the overall data. Importantly for this study, all cases follow patterns similar to that in Fig. 1, albeit on different levels overall (see Table 2 or Fig A1 in the supplementary material).

Sample

At the time of data collection, the 1,192,909 users included in our sample had, on average, 8,028 followers, followed 1,230 users, and had a median of 8,435 posted Tweets. This implies that the users in our sample are somewhat more connected than the full 24-hour sample collected by [22] in September 2022. However, the distribution of these characteristics is skewed, especially concerning the number of followers. While 25% of users have 89 followers or fewer, 68 users have more than 10 million followers. Fig A2 in the supplementary material plots the distribution of these attributes. Only 4% of the users were verified users (47,753 in total). In the robustness checks, we ran the analyses for certain subsets of the sample.

Measuring ethno-religious hostility

We employ the Perspective API of Google Jigsaw's Conversation AI, which is based on convolutional neural networks [58] and is "trained to recognize a variety of attributes (e.g. whether a comment is toxic, threatening, insulting, off-topic, etc.) using millions of examples gathered from several online platforms and reviewed by human annotators" [59]. We applied the API to our Tweets using [57] between December 2022 and February 2023. In particular, we use API's IDENTITY_ATTACK attribute to measure ethno-religious hostility. This attribute was designed to measure "Negative or hateful comments targeting someone because of their identity" [59]. In this context, we assume that identity proxies ethnicity, race, or religion (similar to [42]). In the models below, we use the continuous version of the outcome provided by the Perspective API, which measures from 0 to 100% the probability that a comment is hostile. Averaged over the whole data, the probability that a Tweet is hostile by this definition is 21.6%. In alternative approaches, we ran the same analyses for binary versions of this variable (see robustness checks).

To validate the results obtained through Jigsaw's API, we drew a random sample of about 2,000 Tweets, stratified by case, the before-after period, and the predicted hostility (oversampling Tweets with a predicted probability of 0.75). We then employed ChatGPT 3.5, which has been shown to mirror well the judgment of human coders (usually considered the gold standard) [60, 61] or even exceed it [62]. As proposed by Gilardi and colleagues [60], we developed instructions for ChatGPT and fed each Tweet with the same instruction one by one (see the supplementary material for details). This was done twice for each Tweet in May 2023. The results show that the predictions of both approaches correlate highly ($r=0.8$). As another reliability metric, we calculated the Intra Class Correlation Coefficient, which shows 77% agreement with Perspective and either one GPT prediction. To check whether our approach is also valid for the highest levels of hostility, we categorized hostility probabilities as 0 (<=90%) and 1 (>90%) and calculated F1 scores.[6] This led to an F1 score of 0.9, with a precision of 0.98 and a recall of 0.82. Thus, in all metrics, our approach appears to be reliable. We assume that the two algorithms are largely independent, as they follow a different logic and are mostly based on different training data. This, in combination with the fact that ChatGPT has already performed well with other tasks, gives us the confidence that our machine-learning predictions are well suited to capturing a Tweet's hostility accurately.

Statistical models

We employ Ordinary Least Squares to compare the average value of ethno-religious hostility in the week before and after an attack, as captured by a dummy variable (*attack*). This is a straightforward test of our first hypothesis. However, a simple before/after comparison would not take into account different trends in hostility in the two periods. To test for differences in trends, we add a running variable (*time_since_attack*) counting the minutes until and after an attack—roughly ranging from −10,000 min (7 days before) to +10,000 min (7 days after). We use this variable and interaction term between the *attack* dummy variable and *time_since_attack* variable to perform an interrupted time series analysis. This way, we can test the effect directly after attacks for our first hypothesis and the trend afterward for our second hypothesis. In particular, the underlying equation is:

$$\begin{aligned} p\,(hostility) =& \beta_0 + \beta_1 attack + \beta_2 time\_since\_attack \\ & + \beta_3(attack \times time\_since\_attack) + \beta_4 case + \varepsilon \end{aligned}$$

Where β_1 is the immediate attack effect on hostility (i.e., at time point 0), β_2 captures potential time trends in hostility before an attack (i.e., conditional on attack=0), β_3 models the trend after the attack, and β_4 captures differences across the ten cases. For Hypotheses 3a and 3b, we re-estimate this model for the subset of Tweets that does not mention Muslims or Islam at all.

To examine the generalizability of Hypotheses 1 and 2, we employ models that allow both the attack effect and the trends to vary across cases. This implies a three-way interaction model between the attack dummy, the time variable, and the case variable. The equation of this model is the following (note that the required two-way interaction terms are excluded from the equation for simplicity, but included in the estimated models):

$$\begin{aligned} p\,(hostility) =& \beta_0 + \beta_1 attack + \beta_2 time\_since\_attack + \beta_3 case \\ & + \cdots [twowayinteractionterms] \cdots \\ & + \beta_4(attack \times time\_since\_attack \times case) \\ & + \beta_5 case + \epsilon \end{aligned}$$

[6] The F1 score is the harmonic mean of precision (the proportion of true positive predictions from all positive predictions) and recall (the proportion of true positive predictions from all actual positive observations).

This model is largely analogous to the one above, with β_4 in this equation modeling how the after trend varies across the ten cases. As the interpretation of the parameters in three-way interactions is quite complex, we plot the results below and summarize the main findings. The full table of results is available in the supplementary material. In all models, we cluster standard errors on the user level to account for autocorrelation using [63].

Finally, we run user-level fixed effects models (FE) to analyze changes in the Tweeting behavior *within* users over time, which allows us to test Hypothesis 4 on the mechanisms underlying shifts in hostility levels after attacks. These FE models allow us to inspect whether individual users indeed became more hostile after attacks. FEs offer the strong benefit of automatically controlling for all time constant or slow characteristics on the user level. This does imply, however, analyzing only the subset of 324,519 users who Tweeted both before *and* after an attack—approximately 27% of the users in our overall data set.

Results

The simple comparison of the before and after attack Tweets in Model M1 of Table 3 shows that Tweets are indeed more hostile after the attacks (full model in Table A1 in the supplementary material). In the aggregate, the average probability that a Tweet is hostile is 5.4% points higher after an attack on the 0–100-point scale.

To test whether there are different trends in the before and after attack periods, we perform an interrupted time series analysis (M2 in Table 3). As expected, the results show a negligible and statistically non-significant trend before the attack (beta $=-0.000001$, $p=0.851$), a 9.9% points jump toward hostility at the time of the attack, and a -0.001 trend after the attack. As the running variable is measured in minutes, this implies that the attack effect weakens, on average, by 0.06 points per hour after the attack, or by 1.44 points per day. Consequently, the boost in hostility due

Table 3 Regression models

	M1: Mean comparison	M2: Time trends	M3: Without Tweets about Muslims
Attack	5.414 ***	9.877 ***	8.156 ***
	(5.321–5.507)	(9.721–10.033)	(7.810–8.503)
Time in minutes		0.000	-0.000
		(-0.000–0.000)	(-0.000–0.000)
Attack × time		-0.001 ***	-0.001 ***
		(-0.001 – -0.001)	(-0.001 – -0.001)
Case dummies	✓	✓	✓
Intercept	11.722 ***	12.009 ***	8.276 ***
	(11.429–12.015)	(11.694–12.325)	(7.584–8.967)
Observations	4,596,300	4,596,300	909,841
AIC	39872345.708	39791994.051	7720340.665

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Note M1 compares the average probability that a Tweet is hostile before and after an attack. M2 models trends in the before and after period and estimates the effect immediately after an attack. M3 replicates M2 but only for Tweets that do not mention Muslims or Islam. Models use cluster-robust standard errors at the user level to the attack is expected to have leveled off about seven days after an attack. These findings are in line with both Hypotheses 1 and 2.

Next, we examine how these findings differ across cases by adding a three-way interaction between the attack dummy, the time variable, and the cases (full model shown in Table A1 in the supplementary material). Results of this model reveal some variation in both the effect of the attack itself and the trends. Fig 2 shows that the attack effect ranges from 15.3 points in the case of the Barcelona attack to 6.91 for the Strasbourg attack—the most recent attack in our data. All cases have a negative trend in the after period (from −0.08 points per hour for the Manchester attacks to -0.04 for the Strasbourg attack). This signals that the impact of an attack is relatively short-lived and, in all cases, returns to its initial value a few days after an attack. Differences across cases notwithstanding, the results support Hypotheses 1 and 2.

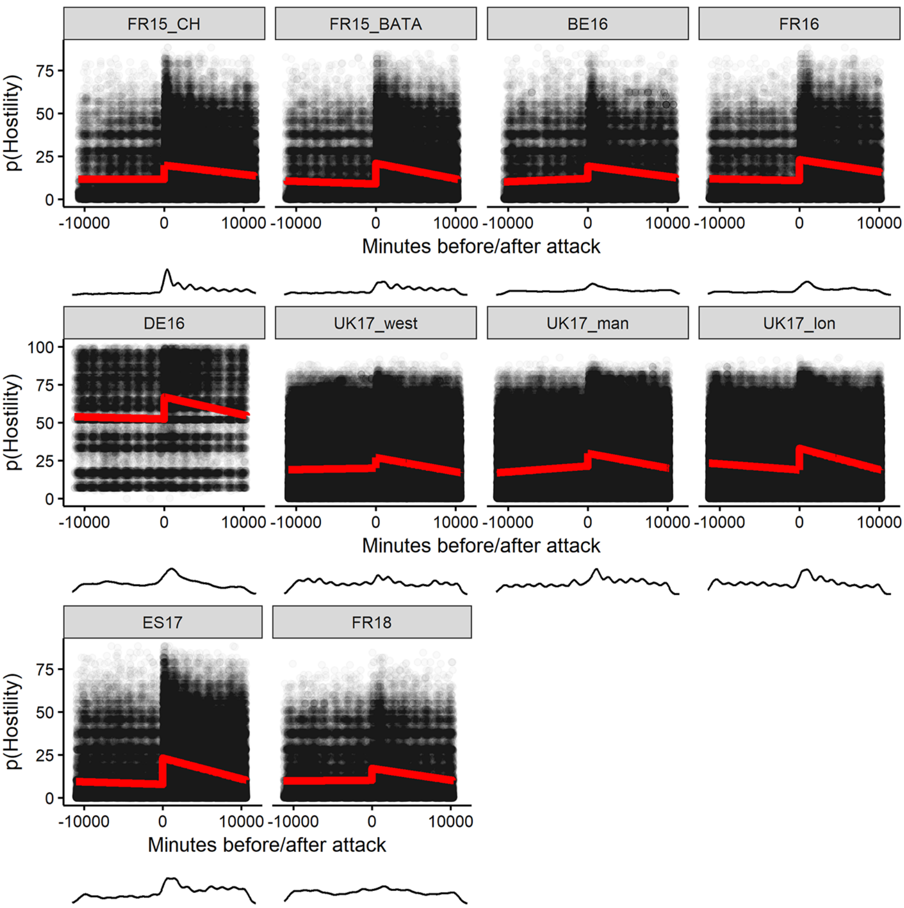

Fig. 2 Trends and changes in ethno-religious hostility before and after terror attacks by case. *Note* The red line represents the average predicted hostility in Tweets over time, with minutes before and after an attack on the x-axis and hostility levels on the y-axis. Zero on the x-axis pertains to the time of the attack. Each circle represents a Tweet; darker shades signify a higher concentration of Tweets. The lines below each case plot illustrate time trends in Tweet density

To test Hypotheses 3a and 3b on the specific and general components of ethno-religious hostility, we ran the models on a subset of Tweets concerned with immigration-related issues, but which do not mention Muslims or Islam at all. Thus, if a Tweet discussed immigration in general, but also mentioned Islam-related terms at some point, it was excluded from this sub-analysis. This leaves about 20% of the Tweets and a new data set of 909,841 Tweets. Given the nature of the events under investigation, it is not surprising that a significant share of Tweets mentions Muslims or Islam-related terms at some point. However, we now turn to the subset of Tweets that do not mention any such terms, but rather discuss more general immigration-related issues. The results are in M3 of Table 3. We find similar, but slightly weaker attack effects when excluding Tweets mentioning Muslims or Islam. In this subset, the hostility rises by about 8.16 points during an attack and levels off after 7.2 days. Hence, while the increase in hostility is strongest for Tweets referring to Muslims and Islam (in line with Hypothesis 3a), we do also observe a substantial and statistically significant effect for the general immigration discourse that does not at all mention Muslims or Islam (in line with Hypothesis 3b).

As a final step, we estimate a user-level fixed effects (FE) model to analyze changes in the hostility of Tweets *within* users. To account for the fact that some users posted during several attacks, we again control for case differences in these models. The FE estimate for the comparison of the average hostility before and after terror attacks shows that such attacks indeed changed the behavior of the same users. The average probability of posting a hostile Tweet is 1.95% points higher after an attack than before (M1 in Table 4). However, this is less than half of the estimated effect of the attack in the pooled model reported above (M1 in Table 3), which amounted to 5.4% points. Although individual users become more hostile in response to attacks, compositional differences (i.e., different users posting before and after attacks) make also a significant contribution to the rise of levels of hostility on Twitter in the aftermath of attacks.

When adding time trends to the user-level fixed effects (FE) model, we observe a 3.6% point rise in hostility immediately after the attack (M2 in Table 4). The fixed effect *within* users again accounts for only part of the overall effect of 9.9% points reported above in the pooled model including time trends (M2 in Table 3). Yet, over-

Table 4 User-level fixed effects models

	M1: Mean comparison	M2: Time trends
Attack	1.955 ***	3.594 ***
	(1.922–1.987)	(3.532–3.656)
Time in minutes		-0.000
		(-0.000–0.000)
Attack × time		-0.000 ***
		(-0.000 – -0.000)
Case dummies	✓	✓
Number of accounts	324,519	324,519
AIC	35772915.766	35760643.719

Note M1 is a within-user comparison of the average probability that a Tweet is hostile before and after an attack. M2 tests for trends in the before and after period and estimates the effect immediately after an attack based on within-user variance only all, the results confirm Hypothesis 4, suggesting that the same users post more hostile Tweets after attacks than before.

Interacting the effect of the attacks with the case variable in the user FE model shows again some interesting heterogeneity. While the Tweeting hostility of users changed most after the German attack with an increase of 3.54 points, it is weakest in the case of the Westminster Bridge attack, with an increase of 1.42% points (see the first model in Table A2 in the supplementary material). However, in all cases, the effect remains positive and statistically significant ($p < 0.001$). For the immediate effect directly after the attack, the estimate is again strongest for Germany with an increase of 5.73% points, and weakest for the Manchester bombing in 2017, with an increase of 2.1 points (all $p < 0.001$ see the second model in Table A2 in the supplementary material).

Robustness checks and plausibility tests

One could argue that whether or not a Tweet is hostile is a categorical characteristic. As an alternative dependent variable, we dichotomized the probability into a dummy variable with values 0 (not hostile) for Tweets with a probability below or equal to 75% and 1 (hostile) for those with a probability above 75%. As an alternative to this, we also created the same variable but with a threshold of 90%. We then regressed these outcomes on the predictors using a binomial link function. The results mirror the findings reported above: while there is a flat, statistically insignificant trend in the before period, ethno-religious hostility spikes at the time of the attack and levels off after. For the 75% threshold version of the dependent variable, the chance that a Tweet is ethno-religiously hostile is 2.9 times higher after the attack. Interestingly, this effect is somewhat smaller for the 90% threshold outcome, with an odds ratio of 2.26 (both models are in Table A3 in the supplementary material). This surprising finding can be explained by the fact that extremely hostile and insulting Tweets are more likely to be instantly deleted by Twitter because they contravene the terms of service of the platform. As a result, extreme levels of hostility are underrepresented in our data. A closer inspection of Fig. 2 corroborates this reasoning, as it demonstrates fewer data points (i.e., lighter areas) at the upper end of each y-axis. Consequently, the estimated effects we report in this study are overly conservative.

Second, we estimated linear trends in the models above, but this may hide some complexity in the temporal dynamics. To account for this possibility, we employ a non-parametric modeling approach to plot trends in the before and after attack periods by case. In particular, we employ generalized additive models with integrated smoothness estimation using restricted marginal likelihood estimation. Fig 3 reveals that the pre-attack period, indicated by the blue lines, shows no distinct trends in either case. Following each attack, the heightened hostility level gradually diminishes, as marked by the red lines in the graph. This trend returns to pre-attack values, typically within a maximum timeframe of 10,000 min (about 7 days). However, in certain instances, such as the 2016 Belgium attack or the 2017 London Bridge attack, we observe a quicker return to pre-attack levels, indicating a more rapid normalization of hostility levels in these cases. Thus, while linear modelling is an approxima-

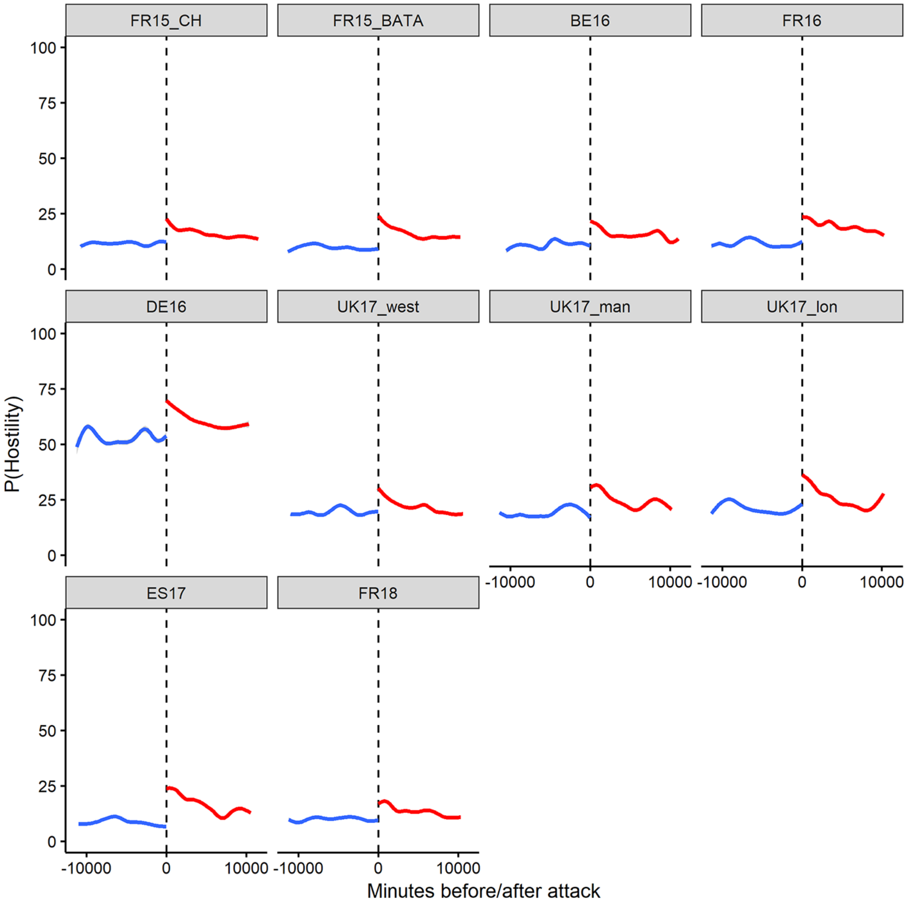

Fig. 3 Nonparametric modeling of trends by case. *Note* The Fig is analogous to Fig. 2, but the trends are estimated nonparametrically as generalized additive models with integrated smoothness estimation, separately for the before and after periods

tion of trends, it appears to capture the underlying structure of our data sufficiently well.

Another major concern with Twitter lies in bots. While unambiguously identifying bots is notoriously difficult, users with a large number of posts have higher rates of bot-like behavior [22 (p. 4)]. To minimize the possibility that our results were mainly caused by bots, we re-ran the analysis but dropped the top 1% of active users in terms of Tweeting.[7] This removes 12,349 highly active users, leaving us with 1,189,935 individual users. Again, the results are highly similar to the ones reported in the main analysis. There is no trend in the before period ($p=0.766$), a 9.9% points jump during the attacks, and a significant negative trend of -1.45 points per day in the time after the attacks (see Table A4 in the supplementary material).

[7] This only relates to Tweets that had not been deleted at the time of the data collection.

Finally, we ran a couple of placebo tests to rule out the possibility that the rise in ethno-religious hostility is simply a matter of chance. As a first placebo, we only looked at the period *before* the attack, as we assumed that nothing relevant to our research question happened during this time. The first placebo is the time point that divides the first week in half. Put differently, we compared trends in the first and last 3.5 days and the change in the middle. As expected, we see a much smaller difference in hostility at this point, of only 1.42% points, and practically no trend in either the first or the second half of this period (the first model in Table A5 in the supplementary material). As a second placebo test, we used the full time frame and generated a treatment at a random time point. In this case, we again find a much smaller difference in hostility between both periods, with a placebo estimate of -1.26 points. We do observe a significant positive trend before and a negative trend in the after period (second model in Table A5). This is to be expected though, as the randomly generated placebo happened to be in the period after the attacks. Thus, the higher levels of hostility after attacks increase the slope of the trend before the random placebo, and the negative trend afterward is part of the after period we show in the main analysis. Together, both placebo attacks reinforce our confidence that the results we report above are causally related to the attacks.

Conclusion

Summary

Jihadist terrorism fuels online hate across cases and countries. We find higher levels of ethno-religious hostility after attacks, but this effect vanishes over time. These findings nicely align with Legewie [11], who used data from the European Social Survey to show that the negative effect of terrorism on immigration attitudes abates over time. In our case, the estimated time until returning to the baseline level is seven days. This also confirms Hopkins [10] and Turkoglu and Chadefaux [38], who both report that the effects of terror attacks are short-lived. We also find statistically significant and sizeable changes in Tweeting hostility *within* users. At first glance, this seems to somewhat contradict research arguing that attitudes remain relatively stable over time [17, 19]. However, this is not necessarily the case, as Kustov et al. [19] primarily argue about long-term changes in immigration attitudes, spanning several years. The present study, in contrast, focuses on short-term changes in expressing hostility on social media. We further find that changes within individuals only account for part of the overall effect. Thus, the overall pattern is shaped by a combination of hostility change within users and changes in the population composition of Tweeters.

Limitations

Firstly, it is unclear how well our findings generalize across social media platforms. This is particularly relevant for Twitter, where users are usually younger, wealthier, and better educated than other Internet users and especially more so than nonusers

[64]. The lack of representativeness of Twitter users complicates inference to other populations.

Secondly, the fact that we used the historical version of the Academic Twitter API leads to the issue of deleted data. Twitter's policy during our period of investigation included the removal of extreme hate speech. The fact that we find weaker effects for more extreme forms of hostility—that is, for Tweets that were 90% hostile vis-à-vis Tweets that were 75% hostile—corroborates this reasoning. As a consequence, the results we present in this study are overly conservative, underestimating the boost in hateful Tweets posted after an attack but then deleted. One might think this concerns earlier attacks more, as more time would have passed in which Tweets could have been deleted. However, the findings also indicate that neither the total level of hostility nor the attack effects are significantly lower for attacks that happened earlier. One explanation for this could be that most extreme hate speech is detected and deleted immediately. In this case, Tweets that have survived for a certain amount of time are likely to remain on the platform. In addition to the deletion of Tweets by Twitter, historic Tweets disappear if they are deleted by the author, or if the author deletes their account. However, it is less likely that this systematically relates to the results presented in this study.

A third limitation is that the classification of Tweets as hostile based on machine learning is something of a black box, due to the proprietary nature of the Perspective API. This is unfortunate, as science should aim for understanding and replicability. However, we also do not know the neural processes that lead to classification decisions in humans. From an applied perspective, the most important issue is that the algorithm performs well, which it does according to our validation.

Finally, it is hard to perfectly identify which users are bots, which means that parts of our data may have been generated by machines instead of humans. We tried to tackle this issue by removing highly active accounts as a robustness check. Moreover, bots — by definition — are programmed to write content on an automated basis (e.g. at a specific hour; see [22]). Given that the terror attacks are exogenous events, such automated posting should, at worst, bias the attack effect toward zero. Yet, the problem of automatic activities might be stronger for more recent attacks, as the number of bots operating on Twitter has increased over time [22]. However, we do not observe differences in trends by year in our data.

Discussion

In sum, our results show that major Jihadist terrorist attacks boost ethno-religious hostility in the digital socioscope Twitter and make individual users more hostile. There is a substantial rise in hostility in the general migration discourse on Twitter after the attacks. Given the importance of online hostility and hate speech for offline behavior, our findings have important implications for the social cohesion of modern multi-ethnic societies. Indeed, online hostility can have real-world consequences, shaping election outcomes [30] as well as violence against refugees [31]. Moreover, because the content of social media is not only influenced by but also influences the content of mass media [28, 29], online hostility may have a spillover effect on traditional and right-wing [65] news outlets. In turn, far-right violence can lead to increased online visibility and influence of right-wing extremists [66]. Finally, social media plays a crucial role as a source of political information for the younger generation. The rise of online ethno-religious hostility, thus, may have an especially pronounced impact on those who are in their impressionable years.

The rise of ethno-religious hostility is likely to become even stronger following the takeover of Twitter (now called X) by Elon Musk, who sees himself as a "free speech absolutist" and purchased the platform with the explicit goal of keeping more controversial content. With this new policy in place, online hatred is likely to become even more widespread — or, at least, will remain more visible. This will likely make the platform (even) more toxic after terrorist attacks. Unfortunately, the abolition of the free academic API by Elon Musk will be a great barrier to future research using Twitter or X data.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42001-024-00272-9.

Acknowledgements An earlier version of the paper was presented at the fourth Conference of the German Academy of Sociology. We thank all discussants as well as the anonymous reviewers for their very helpful comments and feedback.

Funding Christian S. Czymara is supported by the Israel Academy of Sciences and Humanities & Council for Higher Education Excellence Fellowship Program for International Postdoctoral Researchers. Open access funding provided by Tel Aviv University.

Data availability Supplemental materials and code to reproduce the analysis are available on the Open Science Framework at https://osf.io/zdt5b. The data generated during the study are not publicly available owing to privacy issues and Twitter's terms of service. However, corresponding Tweet identifiers are available from the corresponding author upon request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Jacobs, L., & van Spanje, J. (2021). Not all terror is alike: How right-Wing Extremist and Islamist Terror threat affect Anti-immigration Party support. *International Journal of Public Opinion Research*, edaa037. https://doi.org/10.1093/ijpor/edaa037.
2. Kaminski, M. (2015). *All the terrorists are migrants*. *Politico Eu*. https://www.politico.eu/article/viktor-orban-interview-terrorists-migrants-eu-russia-putin-borders-schengen/.

3. Waterfield, B. (2015). *Greece's defence minister threatens to send migrants including jihadists to Western Europe*, https://www.telegraph.co.uk/news/worldnews/islamic-state/11459675/Greeces-defence-minister-threatens-to-send-migrants-including-jihadists-to-Western-Europe.html, *A*.
4. Giani, M., & Merlino, L. P. (2020). Terrorist attacks and minority perceived discrimination. *The British Journal of Sociology*, 1468–4446. https://doi.org/10.1111/1468-4446.12799.
5. Frey, A. (2021). Getting under the skin: The impact of Terrorist attacks on native and immigrant sentiment. *Social Forces*, soab135. https://doi.org/10.1093/sf/soab135.
6. Brouard, S., Vasilopoulos, P., & Foucault, M. (2018). How terrorism affects political attitudes: France in the aftermath of the 2015–2016 attacks. *West European Politics*, *41*(5), 1073–1099. https://doi.org/10.1080/01402382.2018.1429752.
7. Castanho Silva, B. (2018). The (non)impact of the 2015 Paris Terrorist attacks on political attitudes. *Personality and Social Psychology Bulletin*, *44*(6), 838–850. https://doi.org/10.1177/0146167217752118.
8. Böhmelt, T., Bove, V., & Nussio, E. (2020). Can Terrorism abroad Influence Migration attitudes at Home? *American Journal of Political Science*, *64*(3), 437–451. https://doi.org/10.1111/ajps.12494.
9. Finseraas, H., Jakobsson, N., & Kotsadam, A. (2011). Did the murder of Theo Van Gogh Change europeans'. *Immigration Policy Preferences? Kyklos*, *64*(3), 396–409. https://doi.org/10.1111/j.1467-6435.2011.00512.x.
10. Hopkins, D. J. (2010). Politicized places: Explaining where and when immigrants provoke local opposition. *American Political Science Review*, *104*(1), 40–60.
11. Legewie, J. (2013). Terrorist events and attitudes toward immigrants: A natural experiment. *American Journal of Sociology*, *118*(5), 1199–1245. https://doi.org/10.1086/669605.
12. Nägel, C., & Lutter, M. (2020). The Christmas Market Attack in Berlin and attitudes toward refugees: A natural experiment with data from the European Social Survey. *European Journal for Security Research*, *5*(2), 199–221. https://doi.org/10.1007/s41125-020-00066-w.
13. Nussio, E., Bove, V., & Steele, B. (2019). The consequences of terrorism on migration attitudes across Europe. *Political Geography*, *75*, 1–10. https://doi.org/10.1016/j.polgeo.2019.102047.
14. Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior*, *78*, 90–97. https://doi.org/10.1016/j.chb.2017.09.022.
15. Piekut, A. (2021). Survey nonresponse in attitudes towards immigration in Europe. *Journal of Ethnic and Migration Studies*, *47*(5), 27.
16. Tibi, B. (2010). Ethnicity of fear? Islamic Migration and the ethnicization of Islam in Europe. *Studies in Ethnicity and Nationalism*, *10*(1), 126–157. https://doi.org/10.1111/j.1754-9469.2010.01038.x.
17. Kiley, K., & Vaisey, S. (2020). Measuring Stability and Change in Personal Culture using Panel Data. *American Sociological Review*, *85*(3), 477–506.
18. Lersch, P. M. (2023). Change in Personal Culture over the Life Course. *American Sociological Review*, *88*(2), 220–251. https://doi.org/10.1177/00031224231156456.
19. Kustov, A., Laaker, D., & Reller, C. (2021). The Stability of Immigration attitudes: Evidence and implications. *The Journal of Politics*, *83*(4), 1478–1494. https://doi.org/10.1086/715061.
20. Eger, M. A., Mitchell, J., & Hjerm, M. (2022). When I was growing up: The lasting impact of immigrant Presence on native-born American attitudes towards immigrants and immigration. *European Sociological Review*, *38*(2), 169–188. https://doi.org/10.1093/esr/jcab038.
21. Gorodzeisky, A., & Semyonov, M. (2018). Competitive threat and temporal change in anti-immigrant sentiment: Insights from a hierarchical age-period-cohort model. *Social Science Research*, *73*, 31–44.
22. Pfeffer, J., Matter, D., Jaidka, K., Varol, O., Mashhadi, A., Lasser, J., Assenmacher, D., Wu, S., Yang, D., Brantner, C., Romero, D. M., Otterbacher, J., Schwemmer, C., Joseph, K., Garcia, D., & Morstatter, F. (2023). *Just Another Day on Twitter: A Complete 24 Hours of Twitter Data* (arXiv:2301.11429). arXiv. http://arxiv.org/abs/2301.11429.
23. Nunziato, D. C. (2018). The marketplace of ideas online. *Notre Dame L Rev*, *94*, 1519.
24. Menshikova, A., & van Tubergen, F. (2022). What drives anti-immigrant sentiments online? A Novel Approach using Twitter. *European Sociological Review*, *38*(5), 694–706. https://doi.org/10.1093/esr/jcac006.
25. Czymara, C. S., & Dochow, S. (2018). Mass Media and concerns about Immigration in Germany in the 21st Century: Individual-level evidence over 15 years. *European Sociological Review*, *34*(4), 381–401. https://doi.org/10.1093/esr/jcy019.

26. Meltzer, C. E., Eberl, J. M., Theorin, N., Heidenreich, T., Strömbäck, J., Boomgaarden, H. G., & Schemer, C. (2020). Media effects on policy preferences toward free movement: Evidence from five EU member states. *Journal of Ethnic and Migration Studies*, *0*(0), 1–19. https://doi.org/10.1080/1369183X.2020.1778454.
27. Schlueter, E., & Davidov, E. (2013). Contextual sources of perceived group threat: Negative immigration-related news reports, immigrant group size and their interaction, Spain 1996–2007. *European Sociological Review*, *29*(2), 179–191. https://doi.org/10.1093/esr/jcr054.
28. Valenzuela, S., Puente, S., & Flores, P. M. (2017). Comparing Disaster News on Twitter and Television: An Intermedia Agenda setting perspective. *Journal of Broadcasting & Electronic Media*, *61*(4), 615–637. https://doi.org/10.1080/08838151.2017.1344673.
29. Su, Y., & Borah, P. (2019). Who is the agenda setter? Examining the intermedia agenda-setting effect between Twitter and newspapers. *Journal of Information Technology & Politics*, *16*(3), 236–249. https://doi.org/10.1080/19331681.2019.1641451.
30. Giavazzi, F., Iglhaut, F., Lemoli, G., & Rubera, G. (2023). Terrorist Attacks, Cultural Incidents, and the Vote for Radical Parties: Analyzing Text from Twitter. *American Journal of Political Science*, ajps.12764. https://doi.org/10.1111/ajps.12764.
31. Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social Media and hate crime. *Journal of the European Economic Association*, *19*(4), 2131–2167. https://doi.org/10.1093/jeea/jvaa045.
32. Quillian, L. (1995). Prejudice as a response to Perceived Group threat: Population Composition and anti-immigrant and racial prejudice in Europe. *American Sociological Review*, *60*(4), 586.
33. Semyonov, M., Raijman, R., & Gorodzeisky, A. (2006). The rise of anti-foreigner sentiment in European Societies, 1988–2000. *American Sociological Review*, *71*(3), 1988–2000.
34. Blumer, H. (1958). Race prejudice as a sense of group position. *The Pacific Sociological Review*, *1*(1), 3–7.
35. Czymara, C. S., & Schmidt-Catran, A. W. (2017). Refugees Unwelcome? Changes in the Public Acceptance of Immigrants and refugees in Germany in the course of Europe's 'Immigration Crisis'. *European Sociological Review*, *33*(6), 735–751. https://doi.org/10.1093/esr/jcx071.
36. Jungkunz, S., Helbling, M., & Schwemmer, C. (2019). Xenophobia before and after the Paris 2015 attacks: Evidence from a natural experiment. *Ethnicities*, *19*(2), 271–291. https://doi.org/10.1177/1468796818757264.
37. Savelkoul, M., Te Grotenhuis, M., & Scheepers, P. (2022). Has the terrorist attack on Charlie Hebdo fuelled resistance towards Muslim immigrants in Europe? Results from a natural experiment in six European countries. *Acta Sociologica*, *65*(4), 357–373. https://doi.org/10.1177/00016993221088447.
38. Turkoglu, O., & Chadefaux, T. (2022). The effect of terrorist attacks on attitudes and its duration. *Political Science Research and Methods*, 1–10. https://doi.org/10.1017/psrm.2022.2.
39. Oksanen, A., Kaakinen, M., Minkkinen, J., Räsänen, P., Enjolras, B., & Steen-Johnsen, K. (2018). Perceived Societal Fear and Cyberhate after the November 2015 Paris Terrorist attacks. *Terrorism and Political Violence*, *6553*(November 2015), 1–20. https://doi.org/10.1080/09546553.2018.1442329.
40. Álvarez-Benjumea, A. (2020). F. Winter (Ed.), The breakdown of antiracist norms: A natural experiment on hate speech after terrorist attacks. *Proceedings of the National Academy of Sciences* 117 37 22800–22804 https://doi.org/10.1073/pnas.2007977117.
41. Fischer-Preßler, D., Schwemmer, C., & Fischbach, K. (2019). Collective sense-making in times of crisis: Connecting terror management theory with Twitter user reactions to the Berlin terrorist attack. *Computers in Human Behavior*, *100*, 138–151. https://doi.org/10.1016/j.chb.2019.05.012.
42. Czymara, C. S., Dochow-Sondershaus, S., Drouhot, L. G., Simsek, M., & Spörlein, C. (2023). Catalyst of hate? Ethnic insulting on YouTube in the aftermath of terror attacks in France, Germany and the United Kingdom 2014–2017. *Journal of Ethnic and Migration Studies*, *49*(2), 535–553. https://doi.org/10.1080/1369183X.2022.2100552.
43. Jović, M., Šubelj, L., Golob, T., Makarovič, M., Yasseri, T., Krstićev, D. B., Škrbić, S., & Levnajić, Z. (2023). Terrorist attacks sharpen the binary perception of us vs. *Them Scientific Reports*, *13*(1), 12451. https://doi.org/10.1038/s41598-023-39035-3.
44. Pennebaker, J. W. (1993). Social mechanisms of constraint. *Handbook of mental control* (pp. 200–219). Prentice-Hall Inc.
45. Flores, P. M., & Hilbert, M. (2023). Temporal communication dynamics in the aftermath of large-scale upheavals: Do digital footprints reveal a stage model? *Journal of Computational Social Science*, *6*(2), 973–999. https://doi.org/10.1007/s42001-023-00218-7.

46. Garcia, D., & Rimé, B. (n.d.). Collective emotions and Social Resilience in the Digital traces after a Terrorist Attack. *Psychological Science*, *30*(4), 617–628. https://doi.org/10.1177/0956797619831964.
47. Kušen, E., & Strembeck, M. (2021). Evacuate everyone south of that line analyzing structural communication patterns during natural disasters. *Journal of Computational Social Science*, *4*(2), 531–565. https://doi.org/10.1007/s42001-020-00092-7.
48. Helbling, M., & Meierrieks, D. (2020). Terrorism and Migration: An overview. *British Journal of Political Science*, 1–20. https://doi.org/10.1017/S0007123420000587.
49. Meuleman, B., Abts, K., Slootmaeckers, K., & Meeusen, C. (2018). Differentiated threat and the Genesis of Prejudice: Group-Specific antecedents of Homonegativity, Islamophobia, Anti-semitism, and anti-immigrant attitudes. *Social Problems*. https://doi.org/10.1093/socpro/spy002.
50. Gorodzeisky, A., & Semyonov, M. (2019). Unwelcome immigrants: Sources of opposition to different immigrant groups among europeans. *Frontiers in Sociology*, *4*, 1–10. https://doi.org/10.3389/fsoc.2019.00024.
51. Meeusen, C., & Jacobs, L. (2017). Television News Content of Minority groups as an Intergroup Context Indicator of differences between target-specific prejudices. *Mass Communication and Society*, *20*(2), 213–240. https://doi.org/10.1080/15205436.2016.1233438.
52. Godefroidt, A. (2023). How Terrorism does (and does not) affect citizens' political attitudes: A Meta-analysis. *American Journal of Political Science*, *67*(1), 22–38. https://doi.org/10.1111/ajps.12692.
53. Bove, V., & Böhmelt, T. (2016). Does Immigration induce Terrorism? *The Journal of Politics*, *78*(2), 572–588. https://doi.org/10.1086/684679.
54. Flores, R. D. (2017). Do anti-immigrant laws shape public sentiment? A study of Arizona's SB 1070 using Twitter Data. *American Journal of Sociology*, *123*(2), 333–384. https://doi.org/10.1086/692983.
55. Spörlein, C., & Schlueter, E. (2020). Ethnic insults in YouTube comments: Social Contagion and Selection effects during the German Refugee Crisis. *European Sociological Review*, jcaa053. https://doi.org/10.1093/esr/jcaa053.
56. Barrie, C., & Ho, J. C. (2021). academictwitteR: An R package to access the twitter academic research product track v2 API endpoint. *Journal of Open Source Software*, *6*(62), 3272. https://doi.org/10.21105/joss.03272.
57. Votta, F. (2020). *peRspective: Interface to the 'Perspective' API* (0.1.0) [Computer software]. https://github.com/favstats/peRspective.
58. Dixon, L., Li, J., Sorensen, J., Thain, N., & Vasserman, L. (2018). Measuring and mitigating Unintended Bias in text classification. *Proceedings of the 2018 AAAI/ACM Conference on AI Ethics and Society*, *67-73*, https://doi.org/10.1145/3278721.3278729.
59. Google (2023). *About the API*. https://support.perspectiveapi.com/s/about-the-api?language=en_US.
60. Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, *120*(30), e2305016120. https://doi.org/10.1073/pnas.2305016120.
61. Ornstein, J. T., Blasingame, E. N., & Truscott, J. S. (2023). *How to Train Your Stochastic Parrot: Large Language Models for Political Texts*.
62. Törnberg, P. (2023). ChatGPT-4 outperforms experts and crowd workers in Annotating Political Twitter messages with zero-shot learning. *arXiv*. arXiv:2304.06588. http://arxiv.org/abs/2304.06588.
63. Blair, G., Cooper, J., Coppock, A., Humphreys, M., & Sonnet, L. (2024). *estimatr: Fast estimators for design-based inference* [Manual]. https://CRAN.R-project.org/package=estimatr.
64. Blank, G. (2017). The Digital divide among Twitter users and its implications for Social Research. *Social Science Computer Review*, *35*(6), 679–697. https://doi.org/10.1177/0894439316671698.
65. Czymara, C. S. (2024). Real-world developments Predict Immigration News in Right-Wing Media: Evidence from Germany. *Mass Communication and Society*, *27*(1), 50–74. https://doi.org/10.1080/15205436.2023.2240307.
66. Van Der Vegt, I., Mozes, M., Gill, P., & Kleinberg, B. (2021). Online influence, offline violence: Language use on YouTube surrounding the 'Unite the right' rally. *Journal of Computational Social Science*, *4*(1), 333–354. https://doi.org/10.1007/s42001-020-00080-x.