ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Álvarez-Benjumea, Amalia; Winter, Fabian

Working Paper

The breakdown of anti-racist norms: A natural experiment on normative uncertainty after terrorist attacks

Discussion Papers of the Max Planck Institute for Research on Collective Goods, No. 2020/5

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Álvarez-Benjumea, Amalia; Winter, Fabian (2020) : The breakdown of anti-racist norms: A natural experiment on normative uncertainty after terrorist attacks, Discussion Papers of the Max Planck Institute for Research on Collective Goods, No. 2020/5, Max Planck Institute for Research on Collective Goods, Bonn, https://hdl.handle.net/21.11116/0000-0005-A676-B

This Version is available at:
https://hdl.handle.net/10419/229991

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Discussion Papers of the
Max Planck Institute for
Research on Collective Goods
2020/5

The Breakdown of Anti-Racist Norms: A Natural Experiment on Normative Uncertainty after Terrorist Attacks

Amalia Álvarez-Benjumea
Fabian Winter

MAX PLANCK
SOCIETY

Discussion Papers of the
Max Planck Institute
for Research on Collective Goods

2020/5

The Breakdown of Anti-Racist Norms: A Natural Experiment on Normative Uncertainty after Terrorist Attacks

Amalia Álvarez-Benjumea / Fabian Winter

February 2020

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
https://www.coll.mpg.de

The Breakdown of Anti-Racist Norms: A Natural Experiment on Normative Uncertainty after Terrorist Attacks

Amalia Álvarez-Benjumea *[1] and Fabian Winter[1]

[1]Max Planck Institute for Research on Collective Goods

February 11, 2020

Abstract

Terrorist attacks can have profound consequences for the erosion of social norms, yet the causes of this erosion are not well understood. We argue that these attacks create substantial uncertainty about whether norms of civic conversation still hold. Observing breaches of these norms then leads people to express their own anti-immigrant attitudes more readily, as compared to a context where these norms are unambiguous. To test our theory, we examine (i) the impact of terrorist attacks on the level of hate speech against refugees in online discussions, and (ii) how the effect of terrorist attacks depends on the uncertainty about social norms of prejudice expression. To this end, we report on the results of a unique combination of a natural and a lab-in-the-field experiment. We exploit the occurrence of two consecutive Islamist terrorist attacks in Germany, the Würzburg and Ansbach attacks, in July 2016. Hateful comments towards refugees in an experimental online forum, but not towards other minority groups (i.e., gender rights), increased as a result of the attacks. The experiment compares the effect of the terrorist attacks in contexts where a descriptive norm against the use of hate speech is emphasized, i.e., participants observe only neutral or positive comments towards a minority group, to contexts in which the norm is ambiguous because participants observe anti-minority comments. Observing anti-immigrant comments had a considerable impact on our participants' own comments after the attacks, while observing anti-gender-rights comments did not. We end by discussing implications of the findings for the literature on social norms, sociological methods and policy.

*Direct correspondece to alvarezbenjumea@coll.mpg.de

Keywords: *terrorist attacks, prejudice, online hate speech, social norms, normative uncertainty, anomie, natural experiment*

1 Introduction

On 18 July 2016, a 17-year-old armed with an axe attacked passengers on board a train heading to Würzburg in the southern part of Germany. Six days later, on 24 July, another attacker injured several people and killed himself when he detonated a backpack bomb in Ansbach, near Nuremberg, in the first Islamist terrorist suicide attack in Germany. Both attacks were later claimed by the Islamic State (IS).[1] In the first attack, five people were seriously wounded. After the first attack, a video released by IS called the attacker "a soldier of the Islamic State". In the second attack, 15 people were injured and the attacker died. The IS-linked Amaq news agency stated that the attacker was an IS soldier (Europol, 2017). The two consecutive terrorist attacks fuelled an already heated public discussion on German policies on migration issues and the "European refugee crisis".[2] In the aftermath of the events, media coverage focused on the dangers of opening borders, broadening a public debate on these policies.[3]

After terrorist attacks, hate crimes often follow suit (King and Sutton, 2013; Disha et al., 2011; Byers and Jones, 2007). The effect is particularly noticeable when the attacker is characterized as a member of a social or religious minority, as exemplified by the wave of anti-Muslim hate crimes that followed the 9/11 terrorist attacks (Disha et al., 2011; Byers and Jones, 2007; Hanes and Machin, 2014), the increase in violence against refugees linked to Islamist attacks in Germany (Jäckle and König, 2018), or the escalation of racial and religious hate speech on Twitter after a murder committed by Islamist extremists in the United Kingdom (Williams and Burnap, 2015). More generally, formal and informal norms of "civic behavior" seem to erode after such attacks and behavior that was not acceptable before becomes more frequent in the aftermath.

[1]The attacks were known, respectively, as the Würzburg train attack (*Anschlag in einer Regionalbahn bei Würzburg*) and the Ansbach bombing (*Sprengstoffanschlag von Ansbach*). The attacks were the 4th and 5th Islamist attacks directed to civil population in Germany, and the ones with the largest number of injured people to the date. The Ansbach attack remained the incident with the most non-fatal injuries until the 19 of December of the same year when a truck was driven into a Christmas market leaving 12 deaths and 56 injuries.

[2]The so-called "European refugee crisis" is the name given to a period beginning in 2015 when the number of refugees entering the European Union (EU) dramatically increased.

[3]This is documented in news coverage of the terrorist attacks and German political events during July and August, 2016 (e.g., Connolly, 2016; Oltermann, 2016; Hack, 2016).

We will explain the erosion of civic behavior by focusing on the most immediate public reaction to terrorist attacks that can usually be observed in social media: the public expression of prejudice gains traction in online environments (Awan and Zempi, 2017; Burnap et al., 2014). We will refer to this as *hate speech*, which is speech intended to promote hatred on the basis of race, religion, ethnicity, or sexual orientation (Gagliardone et al., 2015). As of now, however, little is known about the mechanisms causing this increase. It is well established through observational studies that terrorist attacks have a profound impact on xenophobic attitudes (Legewie, 2013; Walters et al., 2016; Echebarria-Echabe and Fernández-Guede, 2006; Boomgaarden and de Vreese, 2007). The role of the terrorist attacks as trigger events of xenophobic attitudes has lead many scholars (e.g., Hanes and Machin, 2014; Awan and Zempi, 2017) to assume that the rise in online hate results from the change in attitudes. The widespread attitudinal change argument states that terrorist attacks increase xenophobic attitudes and anti-immigrant sentiment because people perceive terrorist attacks carried out by out-group members as inter-group threats. Similar accounts focus on a negative emotional reactions after the terrorist attacks, which could also lead to an increase in stereotyping towards those perceived as out-group (Vasilopoulos et al., 2018). Both mechanisms lead to an increase in prejudice (Riek et al., 2006), and, so the argument goes, to an increase in hate speech as a direct consequence of the change in attitudes.

We argue that the attitudinal change argument misses a crucial point: hate speech is a communicative act and, as such, it is regulated by social norms. Social norms play a decisive role in containing the public expression of prejudice such as xenophobic attitudes. Previous research shows that people veil their true attitudes strategically when they believe their preferences are not socially accepted (Kuran, 1995). They avoid disclosing their political opinions to those whom they believe to hold opposite views (Cowan and Baldassarri, 2018) and express less racist opinions after overhearing someone else doing so (Blanchard et al., 1994). Since there are very few formal rules in many online contexts, social norms play a crucial role in these domains. Recent evidence suggests that emphasizing a social norm against the use of hateful language (Álvarez-Benjumea and Winter, 2018; Cheng et al., 2017) or informally sanctioning it (Munger, 2017) reduces the incidence of online hate speech. People are less likely to express hate when a social norm against its expression is unequivocally in place and changes in expectations about the norm can translate into changes in behavior. The relation between norms and behavior is often direct and the effect of anti-prejudice norms can be observed even when attitudes remain unchanged.

For example, Paluck (2009) found that decreasing social acceptability of prejudice reduced its public expression without affecting personal beliefs. In a field experiment conducted by Blanchard et al. (1994), hearing only one person either condemn or condone racism led subsequent participants to either to endorse or to oppose racist views.[4]

We contribute to this debate by suggesting, and empirically testing, a complementing mechanism that takes into account the social acceptability of expressing prejudice. A central building block of the suggested mechanism is an increase in normative uncertainty caused by terrorist attacks. They lead to a situation of confusion in which previous norms are challenged and do not seem to apply any longer. These situations of little normative guidance have been described as a state of *anomie* by Durkheim (1897), but in his conceptualization apply to a much broader state of lawlessness. We will therefore interchangeably use the terms *local anomie* or *normative uncertainty*, which should be understood to apply to a more narrow, local set of norms. In this state of local anomie after terrorist attacks, people become more receptive to normative cues. They more readily follow the example of the hate speech of others by expressing their own attitudes more blatantly. We thus argue that increases in hate speech after terrorist attacks not only depend on their impact on individual attitudes, but also on the uncertainty of the norm against hate speech.

To test the validity of our mechanism, we report the results of a unique combination of a natural experiment and a lab-in-the-field experiment. This allows us to test an empirically challenging idea: the local anomie created by the attacks is not directly observable, such that both an increase in negative attitudes and an increase in norm uncertainty could result in the same effect on hate speech. We exploit the occurrence of the two consecutive Islamist terrorist attacks in Würzburg and Ansbach in Germany (the natural experiment), combined with a pre- and post-attacks lab-in-the-field online forum in which we exogenously manipulate a descriptive norm against the use of hate speech.[5] Many previous natural experiments examine the effect of terrorist attacks on attitudes towards specific social groups. However, because terrorist attacks are, by their own nature, unpredictable, more sophisticated experiments are normally not available. Therefore, this opportunity to isolate any attitudinal change from the normative climate is unique.

[4]For a discussion on the relationship between social norms and behavior and how normative perceptions can be more malleable than attitudes, see Tankard and Paluck (2016).

[5]All data are available from the authors upon request.

Proper randomization of the participants in the different treatments is the key feature of our identification strategy. Randomization produces comparable groups by randomly assigning participants to the experimental conditions. The resulting groups are balanced with respect other variables and thus these cannot affect the estimates of the treatment effects (Webster and Sell, 2014; Shadish et al., 2002; Baldassarri and Abascal, 2017). Our lab-in-the-field experiment ensures proper randomization of participants into experimental conditions, which guarantees that, on average, participants will have the same level of –increased– prejudice across all treatments and hence isolates the effect of the perceived strength of the norm induced by our treatments. In our case, this means that any changes between conditions can be attributed to the treatment because other factors that could correlate with the use of hate speech are evenly distributed across conditions. We should stress here that the aim of this paper is not to horse race the effect of social norms against other possible mechanisms, but to show how norms are part of the change in hate speech after the attacks.

The experimental online forum covers discussions on two different social topics: refugees and gender rights including feminism and Lesbian-Gay-Bi-Transgender (LGBT) rights. Participants in the forum are randomized into three different conditions which vary the descriptive norm against hate speech: a mixed, a neutral, and a positive condition. The mixed condition consists of a mix of comments from friendly language to actual transgressions of the anti-hate norm. Because this condition does not signal any specific descriptive norm, it does not reduce the normative uncertainty produced by the terrorist attacks. In the neutral condition, we show only neutral or positive comments to the participants. This biases the perception of how many others use hate speech, thus creating a behavioral regularity that signals the existence of a descriptive norm against hate speech. To further emphasize the descriptive norm, we create a positive condition in which only positive comments about the respective minority group are shown. As said before, this design thus allows us to compare the effect of the terrorist attacks in contexts where a descriptive norm against the use of hate speech is clear to contexts in which the norm is uncertain under the assumption that randomization isolates the effect of social norms from the additional effect of attitudinal change and other factors.

In the next sections, we describe the political context in which the attacks took place, we lay out the theoretical foundations of the proposed mechanism, outline our identification strategy,

and present the results of our analysis. Finally, the implications of the findings for the dynamics of online hate after terrorist attacks and the implications for the literature on social norms of communication are discussed.

1.1 Political Situation of the Attacks

The Würzburg and the Ansbach terrorist attacks can be contextualized in the "European refugee crisis". Unrest in the Middle East, especially in Syria and Iraq, caused a massive displacement of people fleeing war and political instability, pushing large numbers of refugees to the surrounding countries and to Europe. These people included mostly asylum seekers, but the possibility of hostile individuals, including IS militants, also reaching the EU was widely discussed in the media. Across Europe, the mass immigration generated sympathetic responses towards the newcomers, but also precipitated fears related to the capacity of assimilation and fuelled anti-immigration and populist discourses (Andersen et al., 2017). Public opinion linked the crisis to a surge in Islamist terrorism and fed a narrative around refugees as threatening security and western values, a frame frequently reinforced by the media (Greussing and Boomgaarden, 2017).[6] Negative attitudes toward minorities became more common (Wike et al., 2016), and the widespread attitudes of prejudice increased a fear of an upswing in hate crimes.[7] As the European country that welcomed the highest number of refugees, Germany's reaction attracted a lot of attention. Chancellor Angela Merkel pursued an open border policy, but many others in the country challenged these policies, and anti-immigration parties gained support.[8] In addition to the political reactions, public acceptance of immigrants in Germany decreased from 2015 to 2016 (Czymara and Schmidt-Catran, 2017) and violence directed towards refugees has been on the rise (European Union Agency for Fundamental Rights, 2016).

The Würzburg and Ansbach attacks were the Islamist attacks directed to civil population in

[6]The number of suspects arrested for Islamist terrorism has steadily increased from 2012 to 2016 (Europol, 2017). 2016 alone resulted in 13 attacks and 135 people being killed. A predominant explanation in European and German media for attacks akin to the ones described here is the refugees' religious and ethnic background. See, for example, Stürmer et al. (2018), who analyze the 2016 New Year's Eve sexual assaults in Cologne and show how German mass media connected the attacks against women to refugees and their religious and cultural background.

[7]Police records confirm an increase in hate crimes in the period, with documented cases of discriminatory or hateful acts targeting refugees and immigrants (European Union Agency for Fundamental Rights, 2016).

[8]For instance, the anti-immigration party Alternative for Germany (AfD) was founded in 2013 and narrowly missed the minimum threshold of 5 % to become part of the parliament in that year's federal elections. In the federal election in 2017, the votes for this party rose to almost 13%.

Germany with the largest number of injured people to that date. The Ansbach attack remained the incident with the most non-fatal injuries until December of the same year. This fact and the proximity of the two attacks created a lot of discussion, which heavily affected public opinion. After the attacks, rising levels of online hate speech and the fear that they might cause physical attacks became a matter of concern (Müller and Schwarz, 2017). The hashtags #Merkelsommer – *Merkel's summer* – and #Merkelmussweg –*Merkel should go* – were among the most-discussed topics on Twitter, asking for the resignation of chancellor Angela Merkel, along with general messages against refugees, such as the Twitter hashtag #refugeesNOTwelcome (Kreis, 2017) gaining popularity.[9] Recurring political rallies against the chancellor and her policies were held, and many politicians released comments attacking the open-border policy.

2 Terrorist Attacks Trigger Normative Uncertainty

In this section, we sketch two macro-micro-macro mechanisms to explain an increase of hate speech after terrorist attacks. The first mechanism, the *attitudinal change argument*, is representative of a whole class of mechanisms that are all rooted in individual changes of the speaker, be that attitudinal change or, for example, changes in the emotional state. The second framework, the *social norms mechanism*, is a mechanism that may even work when all internal states are unchanged, but the expectations about other people's reactions are altered. Figure 1 illustrates the different causal pathways we have in mind when explaining the link between terrorist attacks and a rise in hate speech.

The left diagram (Terrorist Attacks - Attitudes - Hate Speech) correspond to the *attitudinal change argument*. Terrorist attacks increase xenophobic sentiment towards those groups linked to the perpetrators of the attacks and consequently lead to more hate speech against that group. The connection between terrorist attacks and attitudinal change is empirically well corroborated in previous research (e.g., Legewie, 2013; Walters et al., 2016; Echebarria-Echabe and Fernández-Guede, 2006; Boomgaarden and de Vreese, 2007), while the link from attitudinal change to hate speech is usually only assumed and not directly tested (Hanes and Machin, 2014; Awan and

[9]The first tweet that contained #refugeesnotwelcome appeared on 10 August 2015 (Kreis, 2017). The hashtag has been recurrently used since then in relation to the refugee crisis, including variants such as #rapefugeesnotwelcome after the New Year's Eve sexual assaults in Cologne in 2016.

Zempi, 2017). Akin mechanisms that imply changes in the internal states of the individuals, such as a negative emotional reaction, will also be represented by the left diagram. Our experiment "controls away" the effect of attitudinal change after the attacks as well as the effect other "internal" factors that could affect hate speech via randomization into the experimental conditions, which allows us to focus on the alternative mechanism: opinions that would have been publicly condemned before the attacks are now more likely to be voiced because the norm is unclear. The *social norms mechanism* is thus a contextual mechanism that works regardless of the changes in the internal states of the individuals.

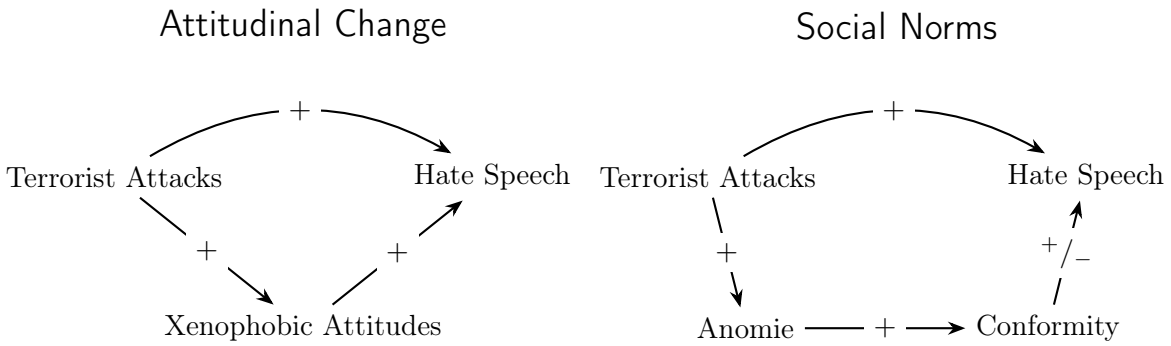


Figure 1: Pathways of the *attitudinal change* mechanism (left) and the proposed *social norms* mechanism (right). The curved arrows represent a macro relationship, the straight arrows the micro mechanism.

Our mechanism is sketched in the right diagram (Terrorist Attacks - Anomie - Conformity - Hate Speech). Terrorist attacks create normative uncertainty and create a situation of confusion or *local anomie*. The previous consensus about the social norm against the public expression of hate erodes and does not provide guidance for behavior. We therefore find ourselves uncertain about what behavior is expected from us, and become more receptive to normative cues. As a result, social conformity increases. The resulting effect on hate speech would depend on the available cues in the social context. Our mechanism thus relies on several elements: hate speech is regulated by social norms, terrorist attacks increase anomie, and anomie promotes the search for cues, particularly the behavior of others. The different building blocks are explained in this section.

Norms regulate hate speech Social norms are shared informal rules that provide the standard of behavior in a social context (Bicchieri, 2006). Their importance for the public expression of hate has been repeatedly established (e.g., Ivarsflaten et al., 2010; Blinder et al., 2013; Crandall

et al., 2002), which makes the expression of prejudice more likely in a private than in a public context (Ford, 2008; Blanchard et al., 1994). The anti-hate norm can be found in the literature under different names, such as egalitarian norm (Crandall et al., 2002), norm of racial tolerance (Weber et al., 2014), anti-racist norm (Ivarsflaten et al., 2010), or anti-prejudice norm (Blinder et al., 2013). The emergence of the norm against hate shapes the way people discuss social issues in public, and also shaped political discourses over the past decades (Mendelberg, 2001).[10] Social norms are inherently ambiguous. They are generally not clearly determined so individuals look for cues in their environment to asses the social acceptability of a behavior like verbalizing hate. A main source of normative information is what others are doing, e.g., their publicly observable behavior (e.g., Tankard and Paluck, 2016; Bicchieri, 2016; Krupka and Weber, 2009).

Descriptive norms serve as cues for normative behavior Behavioral regularities, known as descriptive norms in the literature, are the most available way of perceiving social norms. In past studies, observing social referents, both in media and in real life (Paluck, 2009; Paluck et al., 2016), or simply observing norm-consistent behavior (Blanchard et al., 1994), reduced the expressed prejudice by enhancing the anti-hate norm. Consensus information also changes the perception of the social acceptability of prejudice, leading people to act accordingly. People adjust their reported levels of prejudice after learning the beliefs of the majority (Stangor et al., 2001), and they are more likely to express xenophobic views publicly if they believe others approve of them (Bursztyn et al., 2017). Descriptive norms regulate the decision to use hate speech in online contexts. Users of an online forum are, for instance, less likely to use hateful speech in environments in which a descriptive norm against the use of hate is highlighted (Álvarez-Benjumea and Winter, 2018; Cheng et al., 2017).

Terrorist attacks create local anomie We argue that terrorist attacks challenge the validity of certain anti-hate speech norms. People question whether it is still required to veil one's anti-migrant attitudes or whether it is now permitted to raise them publicly. In reference to (Durkheim, 1897), we refer to this situation of social disorganization in which preexisting social norms not longer work as local anomie. It can be brought about by events that are disruptive,

[10]The dominant norm in the late 1950s in reference to racist speech was one of pro-racist attitudes (Duckitt, 1992). Since then, the public expression of racial prejudice declined over the last decades in western societies (Schuman et al., 1997), and survey respondents were less willing to endorse overt racial prejudices (Huddy and Feldman, 2009).

such as economic crisis, war, or even celebrity suicides (Hoffman and Bearman, 2015). More specifically, events that are perceived as threatening, or are framed as damaging to core values of society, are likely to trigger anomie. The effect increases for heavily publicized events,[11] and obviously terrorist attacks fall in this category.[12] In this situation, individuals seek to regain orientation and thus look at others to form an idea about what is expected.[13] Anomie is not directly observable, but we can, for example, measure the effect of a given event on the perceived strength of the anti-hate norm and how the norm changes in reaction to the event. Events that are exceptional, and therefore disruptive, are likely to cause anomie. For example, after the rather unexpected win of Donald Trump in the 2016 US presidential election, participants in an experiment reported perceiving higher normative acceptability of openly expressing prejudice towards groups targeted during the campaign, but they also adjusted their perception of their own prejudice level by comparing to the new standard (Crandall et al., 2018).

Local anomie increases normative conformity The more uncertain, or anomic, a situation, the more people rely on social cues. Uncertainty facilitates social processes aimed at gaining accurate information about the context, such as self-categorization (Hogg et al., 2007), social comparison (Myers, 1982), and, importantly for our mechanism, normative conformity (Deutsch and Gerard, 1955; Willer et al., 2009). Normative conformity means that people pay more attention to their context in an attempt to learn the appropriate behavior. As a result, people are more likely to copy the behavior of those who are believed to represent the majority (Bicchieri, 2016; Willer et al., 2009).[14] In the case of the expression of prejudice, the uncertainty of the norm has been shown to directly affect the degree of conformity. Zitek and Hebl (2007) found that, as the social norm against hate became less clear, participants were more likely to adjust to a confederate's opinion when reporting their own prejudice. A single person can therefore set a normative expectation in situations of normative uncertainty. Furthermore, normative changes

[11]The level of exposure to media moderates reactions to terrorist attacks, such as increased levels of anxiety (Huddy et al., 2002), in-group favoritism (Traugott et al., 2002), and stereotyping immigrants as a threatening (Boomgaarden and de Vreese, 2007).

[12]In addition to their threatening nature, terrorist attacks are also likely to instill a general sense of uncertainty as to which norms apply by, for example, disclosing previously hidden opinions, or offering a window of opportunity for those individuals who want to frame immigrants as a threat to security to voice their opinion.

[13]Individuals also seek to gain security by reinforcing predominant social categories (Hövermann et al., 2015), therefore triggering prejudice and xenophobic violence (Jäckle and König, 2018).

[14]There is a large empirical literature supporting this idea that, whenever facing an uncertain situation, people will look for behavioral regularities. See the literature on "herding behavior" or cascades (e.g., Banerjee, 1992; Bikhchandani et al., 1992). For a discussion on how behavioral regularities, herding behavior and uncertainty may create the conditions for the emergence of descriptive norms, see Bicchieri (2006, Ch. 6, pp. 216).

in the expression of prejudice are more likely to happen in topics for which the appropriate social norms are unclear (Crandall et al., 2013). It follows that normative uncertainty does not necessarily imply an increase in the expression of hate, but rather it makes people more open to normative influence. Local anomie can thus either amplify or mitigate the escalation of hate speech after terrorist attacks, depending on the social context.

3 Experimental Design: A Combination of a Lab-In-The-Field-Experiment and a Natural Experiment.

Between two waves of data collection in a lab-in-the-field experiment on hate speech in an experimental online discussion forum, two consecutive terrorist attacks took place in Germany. We will occasionally refer to this as the *treatment* in the experimental jargon. We analyze the impact of the terrorist attacks in the hate speech, i.e., speech promoting hatred on the basis of race, religion, ethnicity, or sexual orientation, displayed by our participants. The forum consists of discussions on two different social topics: i.e., gender rights and refugees.[15] The forum experimentally varies the composition of previous comments displayed to the participants, which we will refer to as *experimental conditions*. In the *mixed* condition, which serves as a baseline, no specific norm is signaled. In the *neutral* and *positive* conditions, an anti-hate descriptive norm is highlighted with different strengths. The different experimental conditions offer the opportunity to compare the effects of the attacks in contexts where a descriptive norm against the use of hate speech is highlighted to contexts in which the norm is uncertain. The experimental design includes proper randomization of the participants between conditions, which ensures that other factors are randomly distributed among conditions and isolates the effect of the experimental conditions. The setting also allows us to compare any changes in comments on refugees to changes in comments discussing gender rights to control for common past trends or period effects that could bias the results.

[15]Originally, we selected comments on refugees, LGBT rights and feminism. Feminism and LGBT rights were grouped together in an umbrella category named gender rights to simplify the analysis.

3.1 Design of the Experimental Forum

The forum was designed in three steps: i) selection of topics and pictures, ii) collection and classification of initial comments that would later be shown to the participants and iii) construction of the different experimental conditions. The different steps for constructing the forum and collecting the data are depicted in Figure 2. In the first step, we collected 200 pictures using a list of keywords[16] from different online platforms. Next, we used an online survey in which we asked 90 respondents to decide which topics and pictures would be controversially discussed from the list of 200 pictures. Following that, we selected the top rated pictures and topics in the survey to construct the forum. This forum was available online in a pre-experimental session to collect a first batch of comments on the pictures. A team of three trained raters classified the pool of 840 comments into three categories: neutral, friendly, and hostile. The different experimental conditions consist of a combination of these original comments. By using exactly the same comments before and after the terrorist attack, we avoid the endogeneity problem that often comes with studying peer effects, in which individual and peer behavior are mutually reinforced (Angrist, 2014).

At the beginning of the experiment, participants were randomly assigned to one of the experimental condition and provided with a user name and an avatar. They were told about the experimental nature of the study, but not the actual purpose of the experiment. They were asked to join the conversations and leave comments on the different posts. Once the experiment started, every participant was consecutively presented with the discussions and asked to leave a comment at the bottom of each thread (see Figure 3 for a screenshot of the neutral condition). Importantly, we told the participants that the comments should be readable and on the topic, but intentionally avoided stating any expectations about normative concerns. Also, participants could not see what other participants immediately before them had commented, but only the comments we had previously selected to create the different conditions. This ensures that individual observations are independent, and thus increases the internal validity of our results. Each participant was required to leave a comment on each forum page, with a total of eight comments

[16]The images were obtained from Twitter and Google during March 2016 and we used a set of tags and keywords. Both German and English were used in the search, as both languages are often used on German social media. The following terms and derivatives were used: Sharia, Multiculturalism, Terrorism, Transgender, Gay, Sexism, Discrimination, Refugees, *Aufschrei*, Immigration, Homosexuality, *Einwanderung*, Diversity, Queer, Begging, Atheism, Islamization, Religion, *Tolerist*

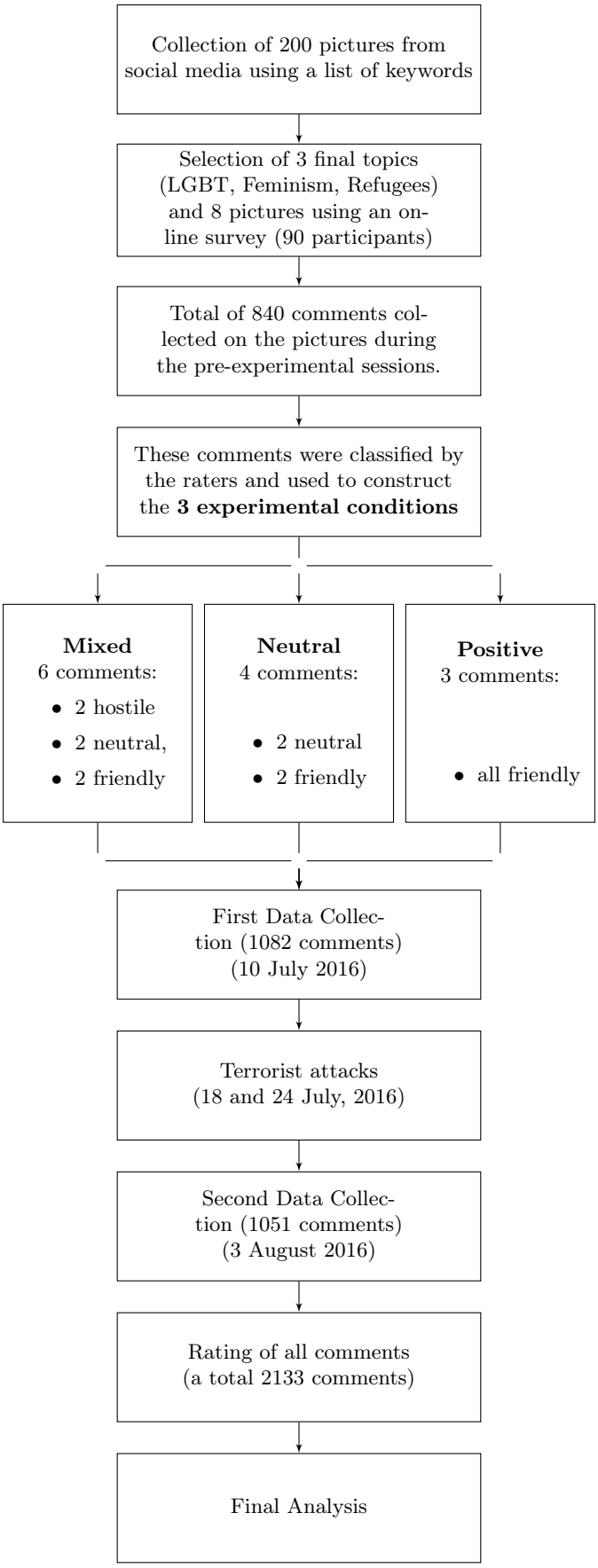


Figure 2: Flowchart of experimental design and online forum set-up

per participant.[17] Each participant gave a comment on each of the eight pictures illustrating the two topics: five pictures on gender rights, and three pictures on refugees. At the end of the experiment, participants received a code that could be exchanged for the payment in an anonymous manner.

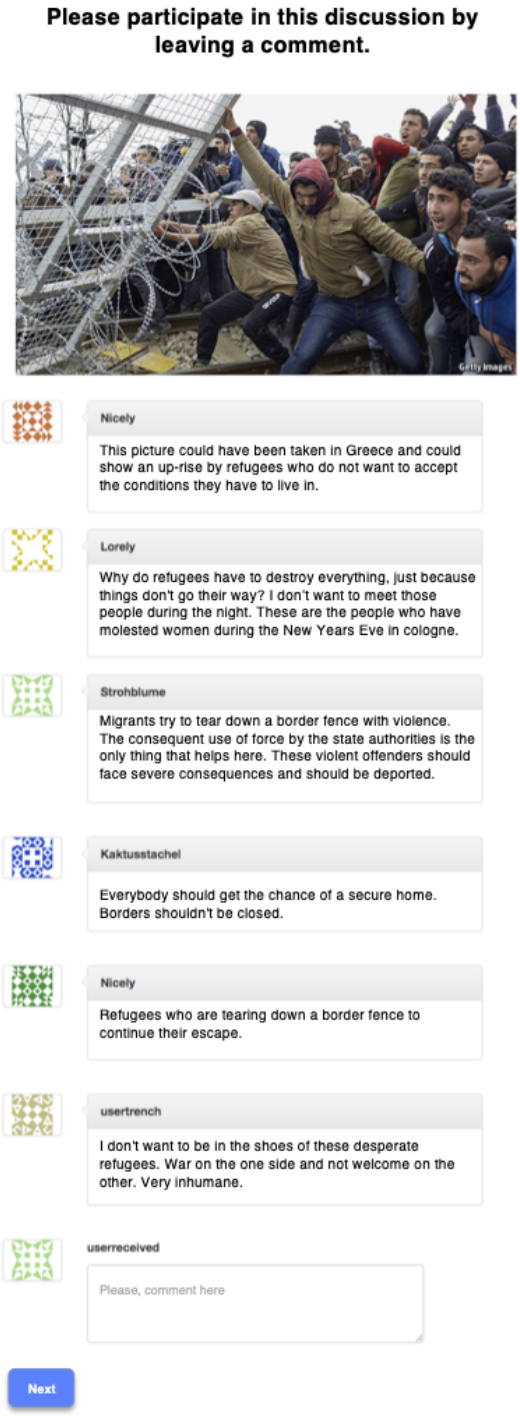


Figure 3: Screenshot of the mixed experimental condition (own translation. In German in the original.)

[17]Commenting was compulsory to go into the next forum page, however, participants could abandon the experiment at any time. On average, each participant left 7.8 comments.

3.2 Experimental Conditions

Participants in the forum were randomized into three different experimental conditions: a mixed, a neutral, and a positive condition. Each condition consisted of a different mix of comments, from friendly language to actual transgressions of the anti-hate norm, which varied the descriptive norm against hate speech. Table 1 shows a summary of the forum content in the different conditions. The *mixed* condition featured a mix of six comments: 2 positive, 2 neutral, and 2 hostile. This configuration did not signal any specific descriptive norm, and therefore did not reduce norm uncertainty. In the *neutral* condition, we showed only four comments: two neutral and two positive comments. This biased the perception of how many others use hate speech and created a behavioral regularity that signals the existence of a descriptive norm against the use of hate in the online forum. The *positive* condition further emphasized the descriptive norm by showing only three positive comments.

Condition	Content
Mixed	6 comments: 2 friendly, 2 neutral, and 2 hostile
Neutral	4 comments: 2 friendly and 2 neutral
Positive	3 comments: all friendly

Table 1: Summary of the content of the online forum in the different experimental conditions

Table 2 shows the number of comments collected by time of data collection (e.g., before or after the terrorist attacks), experimental condition, and topic.

	Before the Attacks		After the Attacks	
Condition	Refugees	Gender Rights	Refugees	Gender Rights
Mixed	135	227	135	228
Neutral	136	225	135	226
Positive	134	225	123	204
Total Comments	405	677	393	658

Table 2: Number of comments (N=2133) per time of data collection (before and after the terrorist attacks), experimental condition, and topic.

4 Research Questions

With the theoretical argument laid out and a thorough description of our experimental methods, we will now present specific research questions. First, we will test whether the terrorist attacks increase the expression of hate speech when no specific norm against hate speech is emphasized. To answer this question we will use data from the *mixed* experimental condition because this condition does not highlight any descriptive norm in particular.

Research Question 1 *Did the terrorist attacks increase the overall level of hate speech in comments about refugees when no specific norm against hate speech is highlighted?*

Second, we test whether the response to the attacks is similar in contexts in which a descriptive norm against the use of hate is highlighted. As discussed above, our experimental approach makes sure that attitudes are randomly distributed and thus on average constant across the experimental conditions. Any remaining differences can thus be attributed to interaction between the anomie created by the terrorist attacks and the normative cues provided by the experimental conditions. If the terrorist attacks generate normative uncertainty, then this uncertainty should be reduced in the experimental conditions with a strong descriptive norm compared to situations where the anti-hate norm is vague (*mixed* condition). Any differences in behavioral patterns among the conditions can be attributed to the anomie mechanism described above.

Research Question 2 *Did the terrorist attacks increase the overall level of hate speech in comments about refugees when a descriptive norm against hate speech is highlighted?*

5 Description of Data, Main Variables and Statistical Approach

5.1 Sample

A total of 139 different participants before the terrorist attacks and 135 after the terrorist attacks for our experiment, as well as 577 raters of the comments, were recruited via a crowdsourcing internet marketplace.[18] The experiment was conducted entirely in German and the sample was

[18] We used the platform www.clickworker.com. It is similar to Amazon MTurk, but with a subsatially bigger work force in Germany.

strictly restricted to residents in Germany. Participants voluntarily registered for the experiment via the platform, which helps us to guarantee that every participant participated in only one condition, either before or after the attack. Unfortunately, the platform has a rather restrictive personal data policy. This lead us to decide not to collect additional data on participants' attitudes, political preferences or individual demographic characteristics. Since we randomize participants into experimental conditions, it shouldn't invalidate our conclusions, but it prevents us from doing any sub-group analyses. We anyways present demographic information on the general characteristics of the workforce for reference in Table 3, which was kindly provided by the platform.

Female	55%
Age	
18-24	28%
25-34	42%
35-44	17%
>45	13%
Employment status	
Student	29%
Employee	26%
Self-employed	15%
Other	20%
N.S	10%

Table 3: Sociodemographic characteristics of the population from where participants were recruited

5.2 Construction of the Mean Hate Score

We define hate speech as speech intended to promote hatred on the basis of race, religion, ethnicity, or sexual orientation (Gagliardone et al., 2015), i.e., hostile expression of prejudice towards minority groups. The outcome of interest is the level of hostility displayed by the participants, which we refer to as the *mean hate score*, and its changes before and after the terrorist attacks, and across the different conditions.

To construct a measure of hate speech, we asked external raters to rate a set of about 30 randomly chosen comments each.[19]. The raters were given the comments in a randomized order and where ignorant of the experimental conditions. Just as the main experiment, the rating task was also completed online using a form provided by us.

Every page of the online form displayed a picture and one comment relating to that picture (see Appendix B). We asked the raters to rate the comment on the following scale: *Is the comment friendly or hostile towards the group represented in the picture? (Give a number from 1 to 9 where 1 means very friendly and 9 means very hostile).* Comments with lower scores, e.g., 1 to 4, are therefore affable, with a cordial language, and often express a positive opinion. On the other side of the spectrum, high scores such as 8 or 9 generally imply abusive language, e.g., "I cannot stand gay people. They should have a psychiatric exam" , or the use of hate terms,[20] e.g., "They can continue walking away from Europe. They are not just war refugees, 90 per cent are nothing but **social parasites** who do whatever they want here" (emphasis added). Appendix A contains examples of comments and their classification.

The continuous *mean hate score* for each comment allows us to study subtle variations, as well as changes in the distribution of the score, and serves as the main variable of interest in our study. To measure the level of agreement of the raters, we used the intraclass correlation (ICC), which yields a value of 0.57. We also computed the Spearman Rank Correlation for random subsamples of the ratings with similar results (see Appendix B). Table 4 gives an overview about the descriptive statistics of the mean hate score.

[19] 498 raters rated 30 comments (86.31%) and 79 rated less than 30 comments. The comments received a median of 7 ratings. Only 36 comments received 2 or fewer ratings and only 22 more than 12, thirteen being the maximum number of ratings a given comment received. The inclusion or exclusion of raters that rated less than 30 comments yields no statistically significant changes in the score, as well as the exclusion of comments with extremely low (< 2) or extremely high (> 12) number of ratings.

[20] We defined hate terms as unambiguously pejorative or derogatory expressions. For a comprehensive compendium of hate terms, see `http://hatebase.org` (last accessed on 7 February 2019).

Table 4: Descriptive Statistics of the Mean Hate Score Before and After the Terrorist Attacks

Experimental Condition	Topic	Time	Mean	Variance	Median	Max.	Min.	1st Quant	3rd Quant	IQR
Mixed	Refugees	Before Attacks	3.90	2.10	3.86	7.57	1.25	2.71	4.87	2.15
Mixed	Refugees	After Attacks	4.46	2.62	4.33	8.12	1.43	3.21	5.54	2.32
Mixed	Gender Rights	Before Attacks	3.12	2.26	2.80	8.60	1.00	2.00	4.06	2.06
Mixed	Gender Rights	After Attacks	3.33	2.87	3.00	8.25	1.00	2.00	4.38	2.38
Neutral	Refugees	Before Attacks	3.96	2.28	3.65	8.60	1.25	2.86	5.00	2.14
Neutral	Refugees	After Attacks	3.98	1.76	3.75	7.43	1.33	3.00	4.86	1.86
Neutral	Gender Rights	Before Attacks	2.97	1.82	2.67	7.00	1.00	1.86	3.86	2.00
Neutral	Gender Rights	After Attacks	3.09	2.02	2.73	7.33	1.00	2.00	4.00	2.00
Positive	Refugees	Before Attacks	3.95	3.12	3.62	8.75	1.00	2.50	5.00	2.50
Positive	Refugees	After Attacks	3.84	2.37	3.50	7.56	1.20	2.65	4.73	2.09
Positive	Gender Rights	Before Attacks	2.85	2.55	2.33	8.88	1.00	1.75	3.44	1.69
Positive	Gender Rights	After Attacks	3.04	2.52	2.43	7.80	1.00	1.89	3.76	1.87

5.3 Statistical Approach

We will analyze the data from our experiment in three steps. First, we measure changes in the level of online hate speech against refugees before and after the attacks and compare it to changes in comments discussing different topics. This allows us to investigate the direct impact of the terrorist attacks on hate speech against refugees. Second, hate speech is compared across the different experimental conditions to test whether the reaction to the terrorist attacks depends on the perceived normative uncertainty. Finally, we will show that this effect is most pronounced for the most hateful comments in our sample.

As described above, each participants gave one comment per picture, and each topic has 2-3 pictures. A comment, i.e., the lowest level observation, is thus nested in participants and topics at the higher level. To accommodate for this nested structure, we estimate the change in the mean hate score in a series of random intercept multilevel linear regression models with two crossed random effects for participants and pictures. We present the analyses in several steps. We first investigate whether the terrorist attacks increased hate speech in comments about refugees when the norm is uncertain (Research question 1). To this end, we analyze only the comments made in the mixed condition. The effect is identified by comparing the level of hate speech before and after the attacks only for comments on refugees, and the interaction effect of the attacks on comments discussing refugees versus the rest of topics (After Attacks $\times$ Refugees, see Equation 1 in the Appendix). To investigate our second research question, i.e., whether people are more receptive towards normative cues in times of anomie, we compare the responses in the mixed condition to the responses in the remaining conditions (see Equation 2 in the Appendix). Third, we establish that this only holds if the norms are really challenged. Since the anti-hate speech norms in comments about gender rights should not be affected by the attacks, we should observe a significant difference-in-difference in the pre- and post-attacks effectiveness of signaling the anti-hate norm for comments on Refugees and Gender rights. Again, we follow a similar strategy and compare the effect of the neutral and positive conditions before and after the terrorist attacks for comments about refugees, and compare these changes to the changes in the Gender rights category (After Attacks $\times$ Refugees $\times$ Experimental Condition, see Equation 3 in the Appendix).

It is important to note that the difference-in-differences is a robust estimator of the effect of the attacks, provided there are no spillovers between the treated and the comparison group. However, previous research shows that social norms are usually susceptible to spillover effects (Keizer et al., 2008; Reno et al., 1993, e.g.,). Accordingly, people exposed to social acceptability of hate speech in one topic could carry this effect over to the rest of the topics. If the spillover effect exists, then the estimated effect of the terrorist attacks on online hate is a conservative estimate.

In the final step, we estimate the differences in the distribution of the mean hate score after the attacks in the different experimental conditions. We do this analysis only for comments discussing refugees. We do this for two reasons. First, to give a more comprehensive picture of the effect of terrorist attacks on hate speech. Second, dividing the hate score into different parts gives us the opportunity to look at changes in the most hateful comments. For each topic and experimental condition, we measure the impact of the terrorist attacks along the conditional distribution of the mean hate score using a quantile regression model (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001).[21] We report results from the 10th to the 95th percentile. Each quantile estimator, in a manner similar to linear regression, minimizes the sum of residuals (Koenker, 2017)(for the specification of the models, see Equation 4 in the Appendix). Rather than predicting the mean change of the hate score, it looks at changes at the quantiles of the score before and after the terrorist attacks.

5.4 Measures to Ensure External Validity

We take several measures in order to ensure the external validity of the experiment. First, our setting allows us to identify the causal effect of the terrorist attacks on hate speech under the assumption that no other potential events could affect the outcome. To our knowledge, there were no terrorist attacks in Germany in the three months leading to the events under study. Finally, we need participants to have been aware of the events, since the reaction to terrorist attacks has been found to be moderated by news exposure (Boomgaarden and de Vreese, 2007). While we do not have direct evidence, we can use Google searches as an indirect indicator of

[21] Quantile regression assumes that "the distribution of the response can be arbitrarily different" (Koenker, 2017, pp.158) conditional in the treatment variable (i.e., the terrorist attacks).

media attention, and the link between the terrorist attacks and the refugee crisis. Figure 4 shows the search interest relative to the highest point on the chart for Germany between 1 June 2016 and 31 August 2016. The search interest should be read as follows: a value of 100 is the peak of popularity for a given search term; a value of 50 means that the term is half as popular; likewise, a score of 0 means the term was less than 1% as popular as the peak. After the terrorist attacks, internet searches in Germany with the keywords *refugee(s)*, *asylum seeker(s)*, and derived terms skyrocketed. Searches that refer to gender rights, namely transgender, feminism, or LGBT, did not show any changes within the period (see Figure A9 in the Appendix).

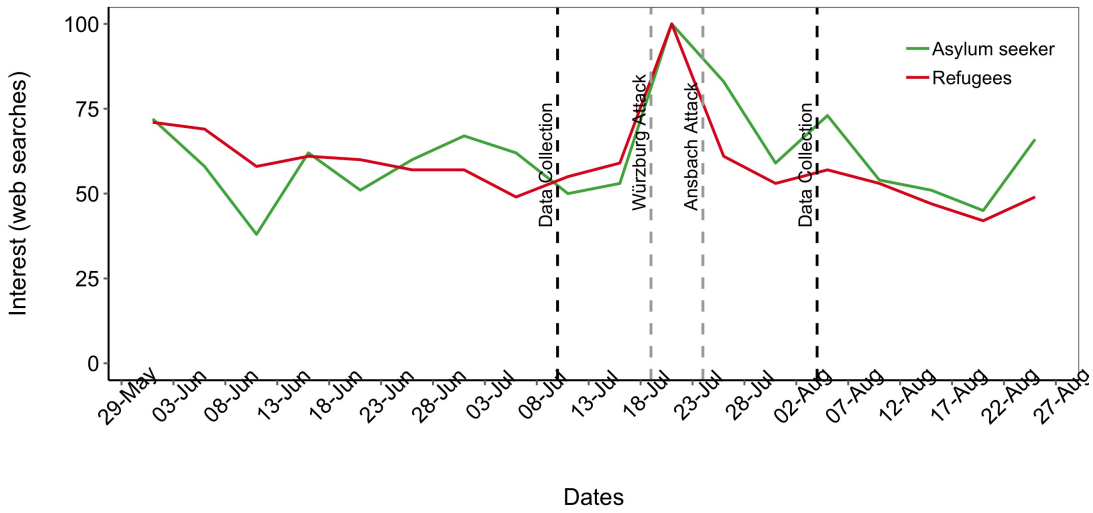


Figure 4: Relative interest (web searches) in Germany for a three months period including the data collection times and the terrorist attacks. The plot shows the popularity of the terms from approximately one month before the first data collection time to a month after the second data collection time. The numbers represent the search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. The dashed lines represent the events of interest: dates of data collection and attacks.

The terrorist attacks and the second instance of data collection are separated by eight days. The first wave of data collection was on the 10 July 2016, nine days before the first terrorist attack. The second wave of data collection started on the 3 August 2016, ten days after the second terrorist attack. While it is difficult to determine the temporal duration of the response to events on theoretical grounds alone, previous findings from survey-based panel studies suggest that the effect of events is long-lived, and persist over an extended period lasting several months (Lecheler and De Vreese, 2011).

Second, our sample of online workers could rise concerns over the participants' motivation to "do a good job in the eyes of their employer", which would lead to an over-estimation of the

effect size. As we have described in the experimental procedure, we took several measures in order to address this limitation: i) participants remained anonymous and no personal data were collected, ii) participation was voluntary, and iii) payment did not depend on performance. We are confident that these design choices created enough detachment between the forum and the marketplace, and therefore we have no reason to assume that the particular treatment effects are qualitatively changed by our sampling strategy.

Third, the use of online convenience samples could raise concerns, even though it is now commonplace in the experimental social sciences. However, it has been shown that online panels are problematic only to the extent that treatment effects (i.e., the reaction to the terrorist attacks) differ between the online sample and the population of interest (Coppock et al., 2018). Since we do not expect our online sample to perceive the environments any differently than the general population, we are reasonably satisfied that there is a sufficient overlap to allow an interpretation of the results of the field experiment. Our identification strategy would also run into problems if the self-selection into the online sample depended on the treatment. For instance, it would be problematic if workers who have more negative attitudes towards refugees would register for a job on the platform with a higher likelihood after the attacks. While we cannot definitely rule this out, it is difficult to think of a reason of why this should be the case.

It is important to note that these concerns are not limited to online workers. Other sampling strategies, such as university students, could also lead to an increased social desirability bias. Using observational data would potentially solve this problem, but the endogeneity associated with it jeopardizes the proper identification of the treatment effect.

6 Results

6.1 Hate Speech towards Refugees, but not towards Gender Rights, Increases after the Terrorist Attacks

First, we analyze the effect of the terrorist attacks in the mixed condition (N=725 comments). As described before, the mixed condition features a mix of comments with different levels of hostility and therefore does not signal any specific descriptive norm. This condition does not

reduce the anomie produced by the terrorist attacks. We expect hateful comments to increase after the attacks in comments in the category Refugees as a result of the increased local anomie in the aftermath of the attacks. Since one is likely experiencing a wide range of reactions in the "real world", the comments in these condition may be the closest to a natural environment.

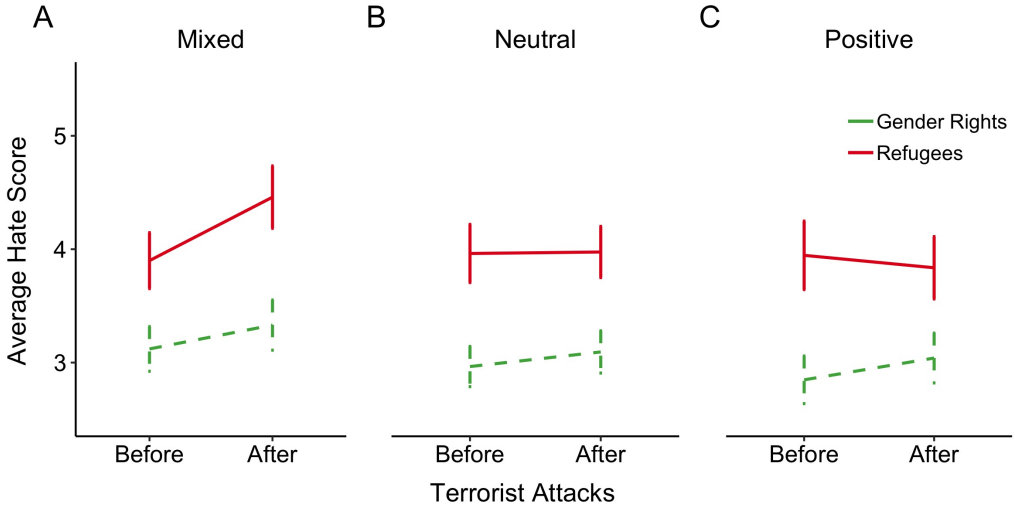


Figure 5: Mean hate score by topic before and after the terrorist attacks. From left to right: Mixed Condition (A), Neutral Condition (B), and Positive Condition (C). The solid red line shows the mean hate score for Refugees, the dashed green line for Gender Rights.

Figure 5A shows an increase in the mean hate score in comments discussing refugees after the terrorist attacks. Model 1 in Table 5 estimates this conditional treatment effect of the terrorist attacks on comments about refugees. The mean hate score increases by 0.56 points following the attacks ($p = 0.004$). To give an intuition about what these coefficients mean, we can look at how changes in the score are translated into changes in hostility in the comments. A change in one score point can be very noticeable when comparing two comments from a thread on the same topic.[22] A comment with a score of 6 reads: "Get rid of the **funny get-up** ...We are in Europe, or more precisely Germany. Whoever wants to live here as to adapt. Multiculturalism, sure, but not that much." A comment in the same picture, but with a score of 7 reads: "That's a very special bird, a **black barn owl** ... " (in reference to a woman wearing a burqa).

[22]Both comments refer to a picture of a woman wearing a burqa sitting in the public transport next to a woman wearing western style clothing. The original comments are in German and translation is our own.

Table 5: Main Results: Regression Estimates and Difference-in-Difference (DiD) Estimates of the Effect of the Terrorist Attacks.

Dependent variable: hate score	Model 1 (Refugees)	Model 2 (All topics)	Model 3 (Refugees)	Model 4 (All topics)
Constant	3.90 (0.19)**	3.12 (0.20)**	3.90 (0.23)**	3.12 (0.20)**
After Attacks	0.56 (0.24)*	0.23 (0.22)	0.56 (0.25)*	0.23 (0.21)
Refugees		0.77 (0.24)**		0.77 (0.25)**
After Attacks × Refugees		0.36 (0.20)†		0.36 (0.19)†
Neutral			0.06 (0.25)	−0.15 (0.21)
Positive			0.05 (0.25)	−0.25 (0.21)
After Attacks × Neutral			−0.55 (0.35)	−0.10 (0.30)
After Attacks × Positive			−0.68 (0.35)†	−0.05 (0.30)
Refugees × Neutral				0.22 (0.19)
Refugees × Positive				0.32 (0.19)†
Refugees × After Attacks x Neutral				−0.48 (0.27)†
Refugees × After Attacks x Positive				−0.67 (0.28)*
Number Comments	270	725	798	2133
Number Participants	90	94	268	274
Number Images	3	8	3	8
$Pseudo - R^2$	.586	.467	.606	.469

Notes: Linear mixed models fit by Maximum Likelihood (ML) with two random effects: participant and image. Fixed Effects Estimates (Top) and Variance-Covariance Estimates (Bottom) for models with mean hate speech score as the dependent variable. Models 1 and 2 include only comments from the mixed condition. Models 3 and 4 include the full sample. Model 1 shows main effects of terrorist attacks for comments in Refugees. Model 2 shows the effect of terrorist attacks for each experimental condition in comments in Refugees. Model 4 shows the DiD estimate for the effect of terrorist attack in each experimental condition and topic. The table lists mean regression coefficient estimates with standard errors in parentheses and p-values calculated based on Satterthwaite's approximations. Significance levels: $^{***}p < 0.000$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.1$, for a two-sided test.

This increase is absent in the category Gender Rights. Model 2 estimates the difference-in-difference between comments on Refugees and Gender Rights before and after the attacks. Before the attacks, comments towards refugees where 0.77 points more negative than those in Gender Rights ($p = 0.029$). After the attacks, comments on Gender rights became 0.23 points more hostile, but this main effect is not statistically significant ($p = 0.28$) and substantially smaller than the 0.56 points increase for post-attacks comments about Refugees. Although the difference-in-difference of 0.36 points is only marginally significant ($p = 0.063$, see the interaction term After Attacks $\times$ Refugees in Table 5), we believe that this analysis supports the finding that the attacks increase the level of hostility against refugees. Also, because this increase is more pronounced in comments discussing refugees when compared to changes in comments discussing gender rights, the treatment effect of the terrorist attacks seems to be different from a mere period effect. To answer our first research question, the terrorist attacks seem to have increased the overall level of hate speech towards refugees when no specific norm was highlighted.

6.2 Descriptive Norms Resolve Anomie After Terrorist Attacks

The central idea of this section is to provide an indirect test of whether the increase in hate speech observed in the mixed environment is caused by the anomie after the terrorist attacks. In this situation, individuals look for normative cues in their environment, such as the behavior of others, to form an idea of what is a socially acceptable response to a situation. In our case, this would translate into a larger effect of the highlighted descriptive norm against hate after the terrorist attacks in comments on Refugees. If this is true, we should expect the increase in hate found in the mixed condition to be reduced or non-existent in the conditions where the norm was highlighted. We should also expect a larger effect of the descriptive norm in comments on Refugees after the attacks when compared to comments on Gender Rights. If, on the contrary, there is no effect of the terrorist attacks on normative uncertainty, we would expect hostility in the forum to increase similarly across all conditions and topics.

To test our idea, we compare the effect of the terrorist attacks on the mixed forum with their effect on the neutral and positive forums for the different topics. Plots B and C in Figure 5 depict the average score before and after the attacks in the neutral forum (B) and the positive forum

(C). There is no visual pre- and post-attacks difference in hate in neither of these conditions in comments about refugees. As before, we construct a multilevel model with a term for the effect of the attacks in the two censored environments compared to the effect of the attacks in the mixed condition. Model 3 in Table 5 again uses only the comments on Refugees and shows that the post-attacks increase of 0.56 points in the mixed condition is offset in both the neutral condition ($\beta = -0.55$, $p = 0.11$) and the positive condition ($\beta = -0.68$, $p = 0.054$). If the participants are confronted with neutral and positive comments only, the stated opinions do not differ before and after the attack ($\beta = 0.014$, $p = 0.95$).[23] This means that the emphasized descriptive norm prevented participants from expressing more hateful opinions after the attacks in the neutral and positive conditions. If we analyze only comments made after the attacks, the mean hate score is significantly smaller in the neutral ($\beta = -0.48$, $p = 0.04$) and the positive condition ($\beta = -0.62$, $p = 0.01$) compared to the mixed condition, which means that the main effects of the experimental conditions became significant after the terrorist attacks. These findings are consistent with our theoretical claim that the increase in hate speech only occurs under anomie and when the descriptive norms are ambiguous.

In Model 4, we added a three-way interaction term (After Attacks $\times$ Refugees $\times$ Experimental conditions) that captures the differential effect of the terrorist attacks in the different conditions by topic. Just as Model 2, Model 4 uses the full sample of comments. Comments towards refugees, in the neutral condition after the attacks, are -0.48 points less hostile ($p = 0.08$), and -0.67 points less hostile in the positive condition ($p = 0.016$) compared to the effect of the experimental conditions after the attacks in the comments discussing gender rights. We thus find (suggestive) evidence for a significantly larger effect of descriptive norms after the attacks in comments on Refugees compared to Gender Rights. These results show that the increase in hate speech in the forum after the terrorist attacks cannot be solely attributed to an increase in negative attitudes towards refugees. An increase in hate resulting from an increase in attitudes would have been consistent across all conditions.

In order to answer our second question, i.e., whether the terrorist attacks increase the overall level of hate speech when a descriptive norm against its use was highlighted, we have looked at

[23]The post-attacks mean hate score when participants are confronted only with positive comments is actually 0.12 (the average hate score is 3.78) points lower than in the pre- attacks mixed condition mean hate score (the average hate score is 4.46), although this is not significant, $p = 0.63$).

the effect of the terrorist attacks in the conditions where the norm was highlighted. In contrast to the post-attacks increase in mean hate score found in the mixed condition, the level of hate speech against Refugees did not change before and after the attacks when the descriptive norm was highlighted. We find that the level of hate, as measured by the mean hate score, did not increase in either of these conditions following the attacks. Furthermore, we find evidence that the effect of the descriptive norms is larger after the attacks when compared to after attacks changes in comments in Gender Rights. We thus conclude that the terrorist attacks did not increase the overall level of hate speech against refugees when a descriptive norm against hate speech was highlighted.

6.3 Descriptive Norms Have the Greatest Effect on Extreme Comments After the Terrorist Attacks

Finally, we show that this increase is driven by a shift of the most extreme comments, i.e., already hostile comments with a rating of 7 and above, and not by moderate comments in the range of 3-5 becoming slightly more negative. If this is true, we should expect the largest treatment effect at the higher quantiles of the distribution because they represent the most hateful comments, which can be regarded as violations of the anti-hate norm. We estimate a quantile regression model.[24] The quantile regression shows how the magnitude of the effect of the terrorist attacks varies across the different percentiles of this distribution. Figure 6 depicts the coefficients of the treatment effect of the attacks on comments about Refugees for the quantiles 0.10 to 0.95 of the distribution of the hate score and the corresponding 95% confidence intervals. The results are shown for all three experimental conditions: mixed, neutral, and positive. Each coefficient corresponds to the change in the τth quantile after the terrorist attacks compared to before the attacks.

[24] We also estimate a linear quantile mixed regression model (Geraci and Bottai, 2014) with participant-specific random intercepts included in the model to account for within-subject dependence. The results using this model are qualitatively similar to the results from the quantile regression we utilize in this paper. The computation of the models with the random term, however, showed several problems of convergence (e.g., converging only at times) due to the small sample size. Therefore the results with the random intercept are not reliable and we decided not to use them.

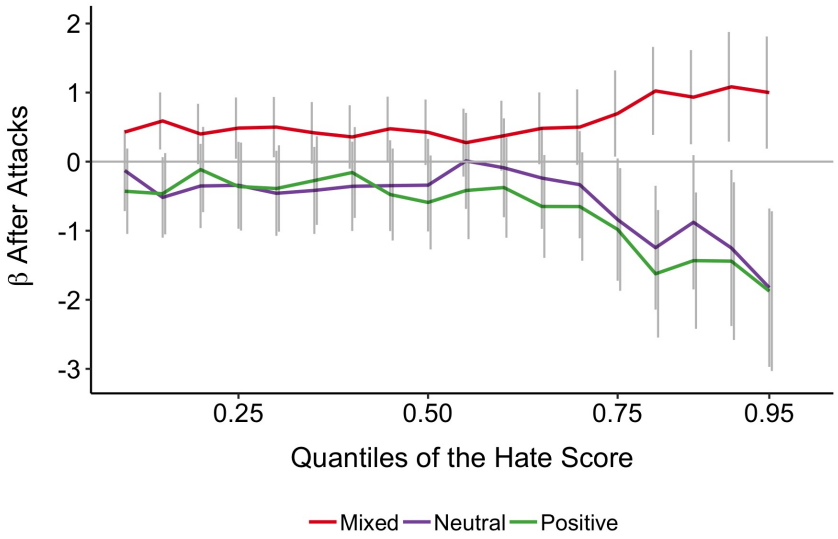


Figure 6: The plots depict the estimated model coefficients of the effect on terrorist attacks for quantiles 0.10 to 0.95 in comments on Refugees for all levels of the descriptive norm: Mixed, Neutral, and Positive. The grey vertical lines represent the confidence interval of the quantile regression coefficients for the effect of terrorist attacks with 95% confidence level.

In all three experimental conditions, the effect of the attacks is more pronounced in the highest quantiles of the distribution of the hate score, which suggests that the average increase after the attacks results from extremely hateful comments, i.e., violations of the anti-hate norm becoming more likely in the mixed condition and less likely in the neutral and positive condition. In the mixed condition, the terrorist attacks increased hate speech for all quantiles of the distribution, but this effect is stronger in the most "hostile" quantiles of the distribution. Compared to an average change of 0.56 points in the score in the mixed condition after the terrorist attacks, the 80% and the 90% percentiles have an increase of 1.023 and 1.083 points, respectively. This effect is both statistically significantly different from 0, an from the other conditions.[25] A similar pattern is observed for the mixed and positive conditions: here, the rating of comments in the 75% quantile and above is at least 1.50 points less negative towards refugees after the attack.

7 Conclusion

The main theoretical argument of our study claims that extreme events, such as terrorist attacks, create normative uncertainty or anomie about which acts are socially permissible. This uncer-

[25]Figure A10 shows the results of the quantile regression for each combination of topic and experimental condition, and the corresponding density plots.

tainty leads people to search for cues in their environment on how to behave, and the behavior of others provides such cues in the form of descriptive social norms. Depending on these social cues, the prevalence of norm violations may increase, stay the same, or even decrease. Our theory thus provides a mechanism that explains that attitudinal change are more likely to materialize in public transgressions of social norms if the norms are challenged by the event.

We apply our reasoning to explain the erosion of norms of civic conversations after terrorist attacks in an online context. A number of studies show the link between terrorist attacks and an increased level of online hate speech (e.g., Awan and Zempi, 2017; Burnap et al., 2014). Our study empirically confirms this finding. We extend the current state of research, which has established a link between terrorist attacks and increased negative attitudes towards the attacker's social group in empirical analyses (Legewie, 2013). Here, we provide additional evidence on how the expression of hate is connected to the uncertainty about social norms. Since anti-hate norms play an important role in containing the expression of prejudice, understanding how terrorist attacks may impact the strength of the social norm is essential to understanding many responses to terrorist attacks.

Our empirical study compares the levels of pre- and post-attacks hate speech in contexts with different strengths of the descriptive anti-hate norm. This allows us to investigate context effects on the comments posted in our experimental online forum and isolate this effect from other possible effects that are averaged out by the randomization procedure. We investigate the effect of the terrorist attacks on normative uncertainty by comparing the effect of descriptive norms after the attacks across all conditions and topics.

We find that the hostility of hate speech increases after Islamist terrorist attacks only towards refugees, and only if previous anti-refugee comments signal the reduced validity of anti-hate norms. In all other conditions, that is, on other topics or when negative comments are absent, hate speech either does not change or is even slightly reduced when only positive comments are shown, even though this reduction is far from statistical significant. The effect of the terrorist attacks on hate speech is thus highly dependent on the local context and the respective norms therein. It also suggests that in natural environments with no external intervention, Islamist terrorist attacks might indeed increase the level of online hate speech against refugees. Our estimation of the absolute effect might be an underestimation of the true effect. We used the

same comments, and therefore the same level of hate speech before and after the attack in our mixed condition. In a natural environment, one would expect that comments would be more negative after the attack, and consequently the reaction to these comments should be even more negative than found in our experiment. This could create a reinforcing loop and thus lead to an escalation of hostility.

Online hate against refugees increases after the attacks both compared with levels of hate against refugees before the attacks, and also relative to the increase towards gender rights. This increase is not found when the anti-hate norm is exogenously manipulated to remain strong. Under the fairly innocent assumption of proper randomization of attitudes into experimental conditions, this difference can be attributed to the interplay between normative context and the normative uncertainty created by the attacks. Our results show that conformity with social norms is greater after the terrorist attacks in topics linked to the events. Our results thus imply that attitudinal changes due to terrorist attacks are more likely to be voiced if the perceived social acceptability of expressing prejudice increases, and that normative uncertainty is an endogenous consequence of terrorist attacks.

The terrorist attacks analyzed here were carried out in a very particular historical context. They happened at the peak of a domestic political crisis due to a sudden influx of refugees, in which immigration laws, integration, and the threat of Islamist terrorism were widely discussed in German media. Does this limit the scope conditions of our theory? While it is impossible to give a definite answer in the absence of additional empirical data, we believe that the proposed mechanism's scope may be much more general than that. It could, for instance, apply to other instances of external shocks, e.g. the surprising election of political candidates, which could alter our beliefs about the expectations of others and rapidly change a previously shared consensus. Bursztyn et al. (2017) show in a donation experiment that monetary transfers to people stating anti-immigrant opinions increased after the election of Donald Trump in 2016. Along these lines, Crandall et al. (2018) suggest that the election also weakened the norm of expressing prejudice for the specific groups targeted by the election campaign. The same could hold for the erosion of anti-hate norms in the public space more generally, particularly if prominent actors publicly break the norms. Our mechanism of increased receptiveness to social cues could also apply to other challenged norms. We would, for instance, predict that an unexpected court ruling

concerning gay marriage or public scandals of underage drinking could create similar patterns of local anomie. Finally, descriptive norms might not be the only cue that could create this effect. Álvarez-Benjumea and Winter (2018) provide suggestive evidence that prescriptive norms could be signaled via instances of sanctioning hate speech. These sanctions could then serve as a cue for normative behavior.

Furthermore, the effect of changing social norms might also vary over time. We have described a mechanism whereby social norms regulate the expression of attitudes; in the long run, however, norms might actually change personal attitudes (Stangor et al., 2001), or our perception of how prejudiced we are in comparison to others (Crandall et al., 2018). Our results support previous empirical findings that show how uncertainty more generally makes people receptive to social influence processes, such as herding behavior (Bikhchandani et al., 1992), conformist influence (Centola et al., 2005), and norm conformity (Deutsch and Gerard, 1955). We thus clarify specific domains where this is the case.

Our findings have direct implications for sociological research methods. Legewie (2013) stresses the importance of dramatic events in one country as a potential source of severe bias in cross-national research. Differences in attitudes between states could be largely driven by the temporal proximity of short-lived attitude changing events. Our findings imply that the particular social context in which the data are collected should also be taken into account because the perceived normativity of certain behaviors in the local context represents an additional source of bias.

It is worth noting that the duration of the effect of the terrorist attacks on normative uncertainty is unclear, and that this effect could change over time. That said, our results suggest that supervisors of public discussions, either in the virtual domain or the real world, may be well advised to implement measures to ensure a well tempered atmosphere. A few bad examples could already lead to an erosion of decency norms. As an unintended consequence, this may lead to the (self-)exclusion of marginalized groups from the discussion, and in the worst case to a breakdown of a whole debate. Naturally, this moderation can be a delicate matter, and the acceptance of these measures may depend to a large extent on existing cultural expectations. It should therefore, if at all, be applied with a strong sense of proportion.

As a last remark, we would like to point out that we obviously did not initially plan to study the

effects of terrorist attacks, but were confronted with it by a tragic coincidence. We acknowledge that a better design to test every path of our theory, and to discriminate between our argument and competing explanations, would have looked much more comprehensive. We could have measured attitudes towards immigrants and gender rights, sociodemographic characteristics of the respondents, media consumption before, between and after the attacks, and maybe anomie. Even more convincing would have been a design in which we would have been able to collect twice the amount of data from the same participants before and after the attacks to estimate individual reactions. But all this would have required us to know about these events in advance, which we obviously did not.

References

Álvarez-Benjumea, Amalia and Fabian Winter. 2018. "Normative Change and Culture of Hate: An Experiment in Online Environments." *European Sociological Review* 34:223–237.

Andersen, Torben M, Giuseppe Bertola, John Driffill, Clemens Fuest, Harold James, Jan-Egbert Sturm, and Branko Uroševic. 2017. "Immigration and the Refugee Crisis – Can Europe Rise to the Challenge?" *EEAG Report on the European Economy* pp. 82–101.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30:98–108.

Awan, Imran and Irene Zempi. 2017. "'I will Blow your Face OFF' – Virtual and Physical World Anti-Muslim Hate Crime." *The British Journal of Criminology* 57:362–380.

Baldassarri, Delia and Maria Abascal. 2017. "Field experiments across the social sciences." *Annual Review of Sociology* 43:41–73.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics* 107:797–817.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Bicchieri, Cristina. 2016. *Norms in The Wild: How to Diagnose, Measure, and Change Social Norms*. New York, NY: Oxford University Press.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100:992–1026.

Blanchard, Fletcher A., Christian S. Crandall, John C. Brigham, and Leigh Ann Vaughn. 1994. "Condemning and Condoning Racism: A Social Context Approach to Interracial Settings." *Journal of Applied Psychology* 79:993.

Blinder, Scott, Robert Ford, and Elisabeth Ivarsflaten. 2013. "The Better Angels of our Na-

ture: How the Antiprejudice Norm Affects Policy and Party Preferences in Great Britain and Germany." *American Journal of Political Science* 57:841–857.

Boomgaarden, Hajo G. and Claes H. de Vreese. 2007. "Dramatic Real-World Events and Public Opinion Dynamics: Media Coverage and its Impact on Public Reactions to an Assassination." *International Journal of Public Opinion Research* 19:354–366.

Burnap, Pete, Matthew L. Williams, Luke Sloan, Omer Rana, William Housley, Adam Edwards, Vincent Knight, Rob Procter, and Alex Voss. 2014. "Tweeting the Terror: Modelling the Social Media Reaction to the Woolwich Terrorist Attack." *Social Network Analysis and Mining* 4:206.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2017. "From Extreme to Mainstream: How Social Norms Unravel." Available at NBER working paper series: `http://www.nber.org/papers/w23415`.

Byers, Bryan D. and James A. Jones. 2007. "The Impact of the Terrorist Attacks of 9/11 on Anti-Islamic Hate Crime." *Journal of Ethnicity in Criminal Justice* 5:43–56.

Centola, Damon, Robb Willer, and Michael Macy. 2005. "The Emperor's Dilemma: A Computational Model of Self-enforcing Norms." *American Journal of Sociology* 110:1009–1040.

Cheng, Justin, Cristian Danescu-Niculescu-Mizil, Jure Leskovec, and Michael Bernstein. 2017. "Anyone Can Become a Troll." *American Scientist* 105:152.

Connolly, Kate. 2016. "Pressure grows on Angela Merkel to start closing Germany's open door." *The Guardian*, July 25. Retrieved June, 2017 (`https://www.theguardian.com/world/2016/jul/25/pressure-grows-on-angela-merkel-to-start-closing-germanys-open-door`).

Coppock, Alexander, Thomas J. Leeper, and Kevin J. Mullinix. 2018. "Generalizability of Heterogeneous Treatment Effect Estimates across Samples." *Proceedings of the National Academy of Sciences* 115:12441–12446.

Cowan, Sarah K and Delia Baldassarri. 2018. ""It Could Turn Ugly": Selective Disclosure of Attitudes in Political Discussion Networks." *Social Networks* 52:1–17.

Crandall, Christian S., A. Eshleman, and L. O'Brien. 2002. "Social Norms and the Expression

and Suppression of Prejudice: The Struggle for Internalization." *Journal of Personality and Social Psychology* 82:359.

Crandall, Christian S., Mark A. Ferguson, and Angela J. Bahns. 2013. "When We See Prejudice: The Normative Window and Social Change." In *Frontiers of Social Psychology. Stereotyping and Prejudice*, edited by Christian S. Crandall and Charles Stangor, chapter 3, pp. 53–69. New York, NY: Psychology Press.

Crandall, Christian S., Jason M. Miller, and Mark H. White. 2018. "Changing Norms Following the 2016 U.S. Presidential Election: The Trump Effect on Prejudice." *Social Psychological and Personality Science* 9:186–192.

Czymara, Christian S. and Alexander W. Schmidt-Catran. 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* 33:735–751.

Deutsch, Morton and Harold B. Gerard. 1955. "A Study of Normative and Informational Social Influences upon Individual Judgment." *The Journal of Abnormal and Social Psychology* 51:629.

Disha, Ilir, James C. Cavendish, and Ryan D. King. 2011. "Historical Events and Spaces of Hate: Hate Crimes against Arabs and Muslims in post-9/11 America." *Social Problems* 58:21–46.

Duckitt, John H. 1992. "Psychology and prejudice: A historical analysis and integrative framework." *American Psychologist* 47:1182.

Durkheim, Emile. 1897. *Suicide: A study in sociology.* New York, NY: Free Press.

Echebarria-Echabe, Agustin and Emilia Fernández-Guede. 2006. "Effects of Terrorism on Attitudes and Ideological Orientation." *European Journal of Social Psychology* 36:259–265.

European Union Agency for Fundamental Rights. 2016. "Current migration situation in the EU: hate crime." Technical report, FRA. Retrieved from `https://fra.europa.eu/en/publication/2016/current-migration-situation-eu-hate-crime`.

Europol. 2017. "EU Terrorism Situation and Trend Report (TE – SAT)." Technical report, Europol, EU, The Hague, Netherlands. Retrieved

from https://www.europol.europa.eu/activities-services/main-reports/eu-terrorism-situation-and-trend-report-te-sat-2017.

Ford, Robert. 2008. "Is racial prejudice declining in Britain?" *The British Journal of Sociology* 59:609–636.

Gagliardone, Iginio, Danit Gal, Thiago Alves, and Gabriela Martinez. 2015. *Countering Online Hate Speech*. UNESCO Publishing.

Geraci, Marco and Matteo Bottai. 2014. "Linear Quantile Mixed Models." *Statistics and computing* 24:461–479.

Greussing, Esther and Hajo G. Boomgaarden. 2017. "Shifting the Refugee Narrative? An Automated Frame Analysis of Europe's 2015 Refugee Crisis." *Journal of Ethnic and Migration Studies* 43:1749–1774.

Hack, Jens. 2016. "German Train Ax–Attack Puts Merkel Migrant Policy back in Spotlight." *The Guardian*, July 18. Retrieved July, 2016 (https://www.reuters.com/article/us-europe-attacks-germany/german-train-ax-attack-puts-merkel-migrant-policy-back-in-spotlight-idUSKCN0ZY2LA).

Hanes, Emma and Stephen Machin. 2014. "Hate Crime in the Wake of Terror Attacks: Evidence from 7/7 and 9/11." *Journal of Contemporary Criminal Justice* 30:247–267.

Hoffman, Mark Anthony and Peter S. Bearman. 2015. "Bringing Anomie back in: Exceptional Events and Excess Suicide." *Sociological Science* 2:186–210.

Hogg, Michael A., David K Sherman, Joel Dierselhuis, Angela T. Maitner, and Graham Moffitt. 2007. "Uncertainty, Entitativity, and Group Identification." *Journal of Experimental Social Psychology* 43:135–142.

Hövermann, Andreas, Steven F Messner, and Andreas Zick. 2015. "Anomie, Marketization, and Prejudice toward Purportedly Unprofitable Groups: Elaborating a Theoretical Approach on Anomie-driven Prejudices." *Acta Sociologica* 58:215–231.

Huddy, Leonie and Stanley Feldman. 2009. "On assessing the political effects of racial prejudice." *Annual Review of Political Science* 12:423–447.

Huddy, Leonie, Nadia Khatib, and Theresa Capelos. 2002. "Trends: Reactions to the Terrorist Attacks of September 11, 2001." *The Public Opinion Quarterly* 66:418–450.

Ivarsflaten, Elisabeth, Scott Blinder, and Robert Ford. 2010. "The Anti-Racism Norm in Western European Immigration Politics: Why we Need to Consider it and How to Measure it." *Journal of Elections, Public Opinion and Parties* 20:421–445.

Jäckle, Sebastian and Pascal D. König. 2018. "Threatening Events and Anti-Refugee Violence: An Empirical Analysis in the Wake of the Refugee Crisis during the Years 2015 and 2016 in Germany." *European Sociological Review* .

Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. "The Spreading of Disorder." *Science* 322:1681–1685.

King, Ryan D. and Gretchen M. Sutton. 2013. "High Times for Hate Crimes: Explaining the Temporal Clustering of Hate-Motivated Offending." *Criminology* 51:871–894.

Koenker, Roger. 2017. "Quantile Regression: 40 Years on." *Annual Review of Economics* 9:155–176.

Koenker, Roger and Gilbert Bassett Jr. 1978. "Regression Quantiles." *Econometrica: Journal of the Econometric Society* pp. 33–50.

Koenker, Roger and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives* 15:143–156.

Kreis, Ramona. 2017. "# refugeesnotwelcome: Anti-refugee Discourse on Twitter." *Discourse & Communication* 11:498–514.

Krupka, Erin and Roberto A. Weber. 2009. "The focusing and informational effects of norms on pro-social behavior." *Journal of Economic Psychology* 30:307–320.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge, Massachusetts: Harvard University Press.

Lecheler, Sophie and Claes H. De Vreese. 2011. "Getting Real: The Duration of Framing Effects." *Journal of Communication* 61:959–983.

Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118:1199–1245.

Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, Oxford: Princeton University Press.

Müller, Karsten and Carlo Schwarz. 2017. "Fanning the Flames of Hate: Social Media and Hate Crime." Available at SSRN: `https://ssrn.com/abstract=2918883`.

Munger, Kevin. 2017. "Tweetment effects on the tweeted: Experimentally reducing racist harassment." *Political Behavior* 39:629–649.

Myers, David G. 1982. "Polarizing Effects of Social Interaction." *Group Decision Making* 14:125–161.

Oltermann, Philip. 2016. "Germany's First Attack by Radicalised Asylum Seeker Alarms Officials." *The Guardian*, July 19. Retrieved June, 2017 (`https://https://www.theguardian.com/world/2016/jul/19/germany-train-attack-could-prompt-rethink-of-counter-terrorism-policy`).

Paluck, Elizabeth Levy. 2009. "Reducing intergroup Prejudice and Conflict Using the Media: A field Experiment in Rwanda." *Journal of Personality and Social Psychology* 96:574.

Paluck, Elizabeth Levy, Hana Shepherd, and Peter M Aronow. 2016. "Changing Climates of Conflict: A Social Network Experiment in 56 Schools." *Proceedings of the National Academy of Sciences* 113:566–571.

Reno, Raymond R., Robert B. Cialdini, and Carl A. Kallgren. 1993. "The Transsituational Influence of Social Norms." *Journal of Personality and Social Psychology* 64:104.

Riek, Blake M., Eric W. Mania, and Samuel L. Gaertner. 2006. "Intergroup Threat and Outgroup Attitudes: A Meta-analytic Review." *Personality and Social Psychology Review* 10:336–353.

Schuman, Howard, Charlotte Steeh, Lawrence Bobo, and Maria Krysan. 1997. *Racial Attitudes in America: Trends and Interpretations*. Cambridge, MA: Harvard University Press.

Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Cengage learning.

Stangor, Charles, Gretchen B. Sechrist, and John T. Jost. 2001. "Changing Racial Beliefs by Providing Consensus Information." *Personality and Social Psychology Bulletin* 27:486–496.

Stürmer, Stefan, Anette Rohmann, Laura Froehlich, and Jolanda van der Noll. 2018. "Muslim Immigration, Critical Events, and the Seeds of Majority Members' Support for Radical Responses: An Interactionist Perspective." *Personality and Social Psychology Bulletin* .

Tankard, Margaret E. and Elizabeth Levy Paluck. 2016. "Norm Perception as a Vehicle for Social Change." *Social Issues and Policy Review* 10:181–211.

Traugott, Michael, Ted Brader, Deborah Coral, Richard Curtin, David Featherman, Robert Groves, Martha Hill, James Jackson, Thomas Juster, Robert Kahn, et al. 2002. "How Americans Responded: A Study of Public Reactions to 9/11/01." *Political Science & Politics* 35:511–516.

Vasilopoulos, Pavlos, George E Marcus, and Martial Foucault. 2018. "Emotional responses to the Charlie Hebdo attacks: Addressing the authoritarianism puzzle." *Political psychology* 39:557–575.

Walters, Mark, Rupert Brown, and Susann Wiedlitzka. 2016. "Causes and Motivations of Hate Crime." Available at SSRN: `https://ssrn.com/abstract=2918883`.

Weber, Christopher R., Howard Lavine, Leonie Huddy, and Christopher M. Federico. 2014. "Placing Racial Stereotypes in Context: Social Desirability and the Politics of Racial Hostility." *American Journal of Political Science* 58:63–78.

Webster, Murray and Jane Sell. 2014. *Laboratory experiments in the social sciences*. Elsevier.

Wike, Richard, Bruce Stokes, and Katie Simmons. 2016. "Europeans Fear Wave of Refugees will Mean more Terrorism, fewer Jobs." *Pew Research Center* 11:2016.

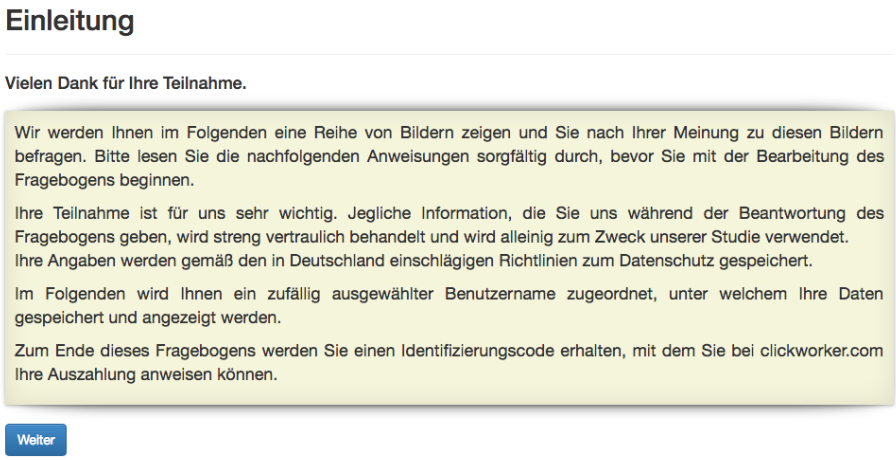
Willer, Robb, Ko Kuwabara, and Michael W. Macy. 2009. "The False Enforcement of Unpopular Norms." *American Journal of Sociology* 115:451–490.

Williams, Matthew L. and Pete Burnap. 2015. "Cyberhate on Social Media in the Aftermath of Woolwich: A Aase Study in Computational Criminology and Big Data." *British Journal of Criminology* 56:211–238.

Zitek, Emily M. and Michelle R. Hebl. 2007. "The Role of Social Norm Clarity in the Influenced Expression of Prejudice over Time." *Journal of Experimental Social Psychology* 43:867–876.

A Appendix A: Materials

A.1 Instructions



Einleitung

Vielen Dank für Ihre Teilnahme.

Wir werden Ihnen im Folgenden eine Reihe von Bildern zeigen und Sie nach Ihrer Meinung zu diesen Bildern befragen. Bitte lesen Sie die nachfolgenden Anweisungen sorgfältig durch, bevor Sie mit der Bearbeitung des Fragebogens beginnen.

Ihre Teilnahme ist für uns sehr wichtig. Jegliche Information, die Sie uns während der Beantwortung des Fragebogens geben, wird streng vertraulich behandelt und wird alleinig zum Zweck unserer Studie verwendet. Ihre Angaben werden gemäß den in Deutschland einschlägigen Richtlinien zum Datenschutz gespeichert.

Im Folgenden wird Ihnen ein zufällig ausgewählter Benutzername zugeordnet, unter welchem Ihre Daten gespeichert und angezeigt werden.

Zum Ende dieses Fragebogens werden Sie einen Identifizierungscode erhalten, mit dem Sie bei clickworker.com Ihre Auszahlung anweisen können.

Weiter

Figure A1: Screenshot of the introduction page of the experiment in German. A translation of the instructions can be found below.

Introduction (In English)

Thank you for your participation.

We will show you a series of pictures and ask you comment on them. Please read the following instructions carefully before you begin the task. Your participation is very important to us. Any information you provide to us during the task will be strictly confidential and will be used solely for the purpose of our study. Your data will be stored in accordance with the relevant data protection guidelines in Germany. You will be assigned a random user name, and your input will be stored and displayed under this username. At the end, you will be given an identification code, which will allow you to claim your payment at `clickworker.com`

Teil 2

Im Folgenden wird Ihnen eine Reihe von Bildern gezeigt, an die sich jeweils eine Diskussion anschließt. Ihre Aufgabe ist es, ebenfalls einen Kommentar in dieser Diskussion zu verfassen.

Ihnen wird für die Dauer dieser Aufgabe ein Nutzername und ein Nutzersymbol zugeordnet, die zur Ihrer Identifikation während der Diskussion dienen. Andere Teilnehmer können so auf Ihre Kommentare reagieren. Sowohl der Nutzername als auch das Symbol können allerdings nicht mit Ihrer realen Identität in Zusammenhang gebracht werden, so dass Sie anonym bleiben.

Ihr Kommentar sollte aus mindestens zwei bis drei Sätzen bestehen. Diese Sätze sollten aussagekräftig sein und sich auf die Bild bzw. die Diskussion beziehen.

User1:"Ich weiß, dass viele Leute Grafittis mögen und sie sogar als Kunst betrachten. Allerdings kann ich Grafittis gar nicht leiden und denke, dass sie die Städte verschandeln."

User2: Das denke ich auch. Die Stadtverwaltung sollte sowas entfernen lassen.

User3: Einige dieser "Grafittis" werden noch in Museen zu sehen sein. Sie repräsentieren die wirkliche moderne Kunst. Das sollte jeder verstehen.

Bitte hinterlassen Sie Ihren Kommentar.

Ein aussagekräftiger Kommentar zu dem oben gezeigten Bild wäre zum Beispiel:

> „Ich weiß, dass einige Leute das schön finden, aber für mich ist es bloße Schmiererei! Es verschandelt die Städte. Die Politik sollte endlich etwas dagegen unternehmen."

Dies hier wäre auch in Ordnung:

> „Ich verstehe die Meinung im ersten Kommentar, aber ich stimme dem nicht zu. Ich finde das schön. Es gehört doch heute einfach mit dazu. Ausserdem sollten junge Leute auch ihre Freiräume haben um sich auszuprobieren."

Der folgende Kommentar hingegen würde nicht als zulässig eingestuft werden:

> "Der schnelle braune Fuchs springt über den faulen Hund. Der schnelle braune Fuchs springt über den faulen Hund. Der schnelle braune Fuchs springt über den faulen Hund."

Ebenso wäre folgender Kommentar kein zulässiger Kommentar:

> "Verlassener Ort!"

Jede Seite wird nur einmal angezeigt. Nachdem Sie Ihren Kommentar abgegeben haben, können Sie zur nächsten Seite wechseln. Sie können jedoch nicht zurückgehen oder vorherige Kommentare bearbeiten.

Bitte drücken Sie "Weiter", wenn Sie bereit sind mit dem Fragebogen zu beginnen. Vielen Danke!

Weiter

Figure A2: Screenshot of the instructions for the experiment in German. A translation of the instructions can be found below.

Instructions (originally in German)

You will see a series of pictures with a discussion below. Your task is to join the discussion on the topic(s) depicted in the picture(s). Please write at least two to three sentences per discussion. These sentences should be meaningful and relate to the picture/discussion.

A valid comment on the discussion above would be:

"I know that some people like them and even consider them to be art. However, I really dislike graffiti or "street art", as some call it. I think it impoverishes the way a city looks."

"I do understand the opinion in comment 1, although I pretty much disagree. Most of the places that

are now covered by graffiti were previously abandoned and looked very dirty and ugly already."

The following comment would not count as valid:

"The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog."

The following is not sufficient as a comment either:

"Abandoned place."

Each page will be shown just once. Once you have finished with your comment you can go to the next page, but you cannot go back or edit previous comments.

Please press the "start" button once you are ready to start the survey.

Thanks!

A.2 Screenshots of the different Experimental Conditions of the Forum

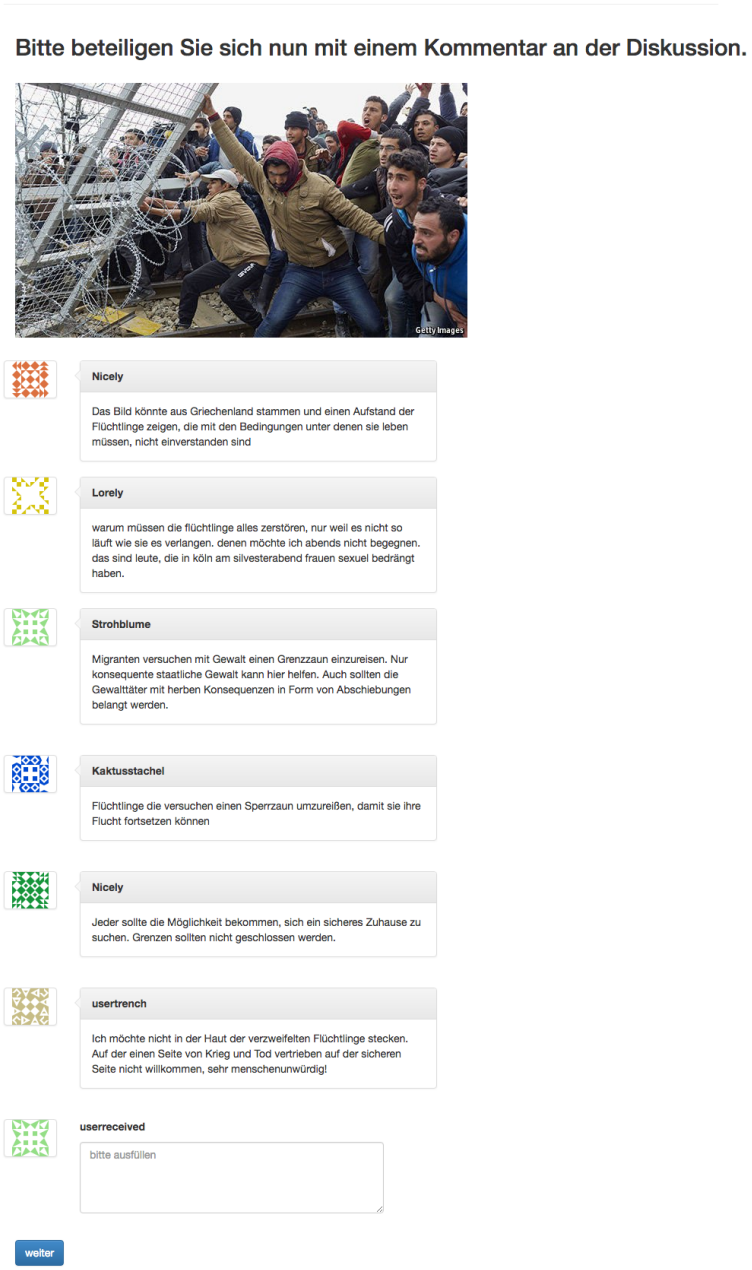


Figure A3: Screenshot of the discussion forum in the mixed condition in German. Examples of the comments in English can be found below.

Figure A4: Screenshot of the discussion forum in the neutral condition in German. Examples of the comments in English can be found below.

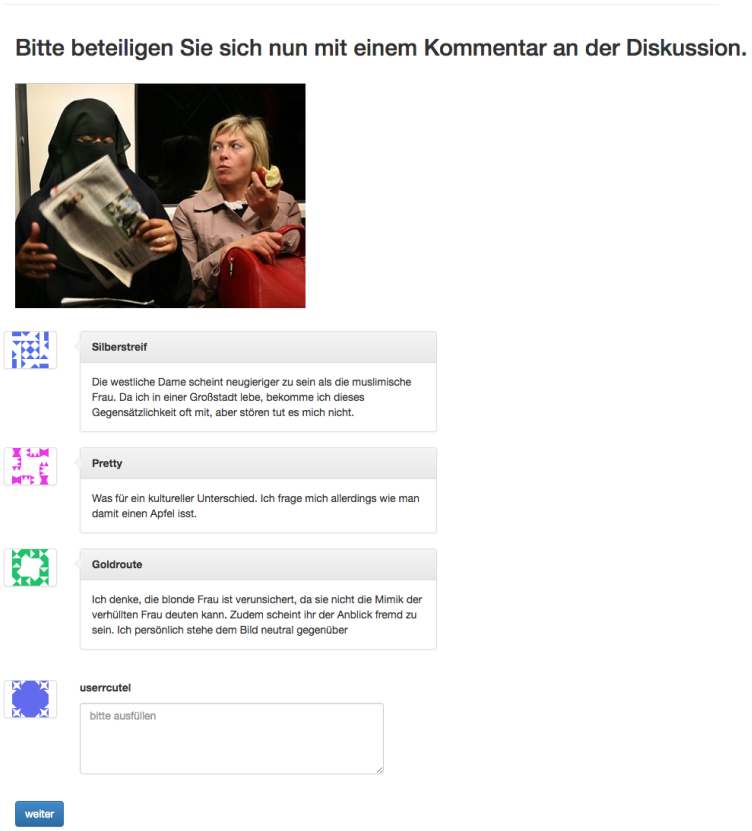


Figure A5: Screenshot of the discussion forum in the positive condition in German. Examples of the comments in English can be found below.

A.3 Screenshot of the Rating Form

Figure A6: Screenshot of the rating form used to rate the comments in German. A translation of the form can be found below.

(Translation of the rating form)

Is this comment friendly or hostile towards the group depicted in the picture? (from 1 to 9 where 1 is extremely friendly and 9 is extremely hostile)

- ☐ It does not apply

Which items are found in the comments?

- ☐ Contains negative stereotypes.
- ☐ Uses racial slurs.
- ☐ Contains words that are insulting, belittling, or diminishing.
- ☐ Calls for violence, threat, or discrimination.

- ☐ Uses sexual slurs.
- ☐ Sexual orientation/gender used to ridicule or stigmatize.

A.4 Examples of Comments

Negative Comments

- Comment1029, mean hate score 8: "ich kann Schwule nicht ausstehen. die sollten sich psychiatrisch untersuchen lassen." *(I cannot stand a gay people. They should have a psychiatric exam.)*
- Comment 391, mean hate score 8.33: "Randalierende Schmarotzer machen alles kaputt. Sie sollen wieder zurück gehen." *(Rioting parasites destroy everything. They should go back again.)*
- Comment 1034,mean hate score 8.66: "Ekelhafte Zurschaustellung der eigenen Perversität. Ich hasse das." *(Disgusting display of one's own perversity. I hate that.)*
- Comment 1261, mean hate score 8.66: "Flüchtlingskrise. Die können weiter wandern weg von europa. das sind nicht nur kriegsflüchtlinge, zu 90 prozent sind es nur sozialschmarotzer die hier machen können was die wollen."*(Refugee crisis. They can continue walking away from Europe. They are not just war refugees, 90 per cent are nothing but social parasites who can do whatever they want here.)*
- Comment 1029, mean hate score 9: "Schwule Kerle sind das Letzte, was ich dulden würde. Schon gar nicht in der Öffentlichkeit." *(Faggots are the last thing I would tolerate, especially in public.)*
- Comment 70, mean hate score 9: "Mit brutaler Gewalt wird versucht, das Sozialparadies Deutschland zu erreichen. Möge der Zaun ewig halten!" *(Brutal violence tries to reach the social paradise of Germany. May the fence last forever!)*

Neutral Comments

- Comment 641, mean hate score 5: "Leider ist dieses Thema noch hoch verpönt, vor allen Dingen wird es sicher nicht der kleine Junge im Leben leicht haben." *(Unfortunately, this topic is still highly frowned upon, above all it will certainly not be easy for the little boy in life.)*
- Comment 892, mean hate score 5: "Generell bin ich dagegen sich in der Öffentlichkeit wild zu küssen. Aber gegen einen Kuss habe ich nichts."*(In general I am against passionate kissing in public. But I have nothing against a kiss.)*

Positive Comments

- Comment 1572, mean hate score 1.66: "Das ist wirklich eine wunderschöne Szene. Es sieht nach einer absolut glücken Familie aus. Wahrscheinlich sind sie glücklicher als so manches hetero Paar."*(This is really a wonderful scene. It looks like an absolutely happy family. They are probably happier than many heterosexual couples.)*
- Comment 1043, mean hate score 2: "Ich finde es gut das gleichgeschlechtliche paare sich immer mehr trauen dies in der Öffentlichkeit zu zeigen. Dies dann auch noch mit einer Bayrischen Tracht zu machen, einfach grossartig."*(I think it is good that same-sex couples are increasingly daring to show this in public. To do this in traditional Bavarian clothes, just great.)*
- Comment 1082, mean hate score 2: "Wirklich ein zuckersüsses Bild! Eine kleine Familie und das Kind sieht so glücklich aus!" *(Truly a sweet picture! A little family and the kid looks so happy!)*
- Comment 469, mean hate score 3: "Zwei Männer in vermutlich einer Lebenspartnerschaft mit einem Kind. Ich finde es toll. Auch finde ich es wichtig, dass wir unsere Kinder heute gleich so erziehen, dass das alles auch normal ist. Und nicht auf dem alten Gedanken bleiben, dass eine Familie nur aus Mama, Papa und Kind(er) besteht, sondern es auch gleichgeschlechtliche "Eltern" gibt." *(Two men presumably in a civil partnership with a child. I like it. Also, I think it is important that we raise our children nowadays with the message that this is normal. And not stick to the odd view that a family consists only of mum, dad and children, but that there are also same-sex parents.)*

- □ Comment 120, mean hate score 2: "Es handelt sich auf dem Bild offensichtlich um ein Menschenmasse von Flüchtlingen, welche auf dem Weg nach Europa sind, ich habe vollstes Verständnis für diese Menschen. Diese Menschen können nichts für den Krieg, der in ihren Ländern inszeniert wird, sie haben alles verloren und wollen (über)leben. Und das ist ihr (Menschen)recht !!!" *(The picture obviously shows a crowd of refugees on their way to Europe, I fully understand these people. There people are not responsible for the war that is staged in their countries, they have lost everything and want to live (survive). And that is their (human) right!!!*
- □ Comment 257, mean hate score 3: "Ich sehe hier eine junge Transfrau. Schön, mutig und selbstbewusst. Sie ist klasse." *(I see a young trans woman here. Nice, brave and self-confident. She is great.)*

B Appendix B: Further Analyses

B.1 Analysis of the Ratings

Figure A7 below shows the ratings per comment. The number of ratings per comment ranges from 1 to 13, with most comments rated between 5 and 10 times each.

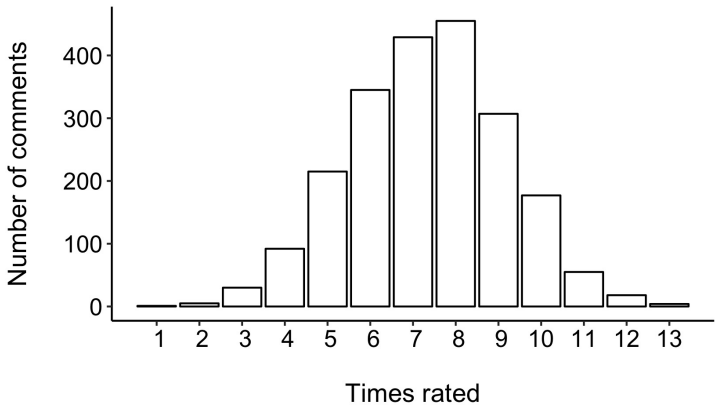


Figure A7: Number of ratings per comment.

Raters were asked to rate a total amount of 30 comments. The rate of attrition of the raters is relatively low, with only 13.7% of them abandoning the task before completion. Table A1 shows the percentage of raters that rated each number of times.

We then analyze the level of agreement of the ratings using both intraclass correlation (ICC), which measures similarity between the ratings of the same comment, and Spearman's rank correlation test, which tests the statistical dependence between the rankings (i.e., the 1 to 9 scale). The ICC coefficient of the whole sample of ratings is 0.57, which is normally understood as a fair level of inter-rater agreement. We conducted robustness checks using a Spearman's rank correlation in different subsets of the comments selected randomly from the whole dataset. The results are shown in Figure A8. The median Spearman's rho coefficient is 0.48, with a maximum value of 0.52.[26]

[26]When comments with very few or too many ratings are not included in the analysis, the Spearman's rho coefficient increases.

Number of ratings by rater	Freq.	Percent	Cum.
1	15	2.60	2.60
2	9	1.56	4.16
3	11	1.91	6.07
4	6	1.04	7.11
5	1	0.17	7.28
6	3	0.52	7.80
7	1	0.17	7.97
9	2	0.35	8.32
10	3	0.52	8.84
11	1	0.17	9.01
12	2	0.35	9.36
13	3	0.52	9.88
14	3	0.52	10.40
15	1	0.17	10.57
16	2	0.35	10.92
18	1	0.17	11.09
20	2	0.35	11.44
22	3	0.52	11.96
24	1	0.17	12.13
25	2	0.35	12.48
26	2	0.35	12.82
27	1	0.17	13.00
28	1	0.17	13.17
29	3	0.52	13.69
30	498	86.31	100.00
Total	577	100.00	

Table A1: Number of ratings per rater

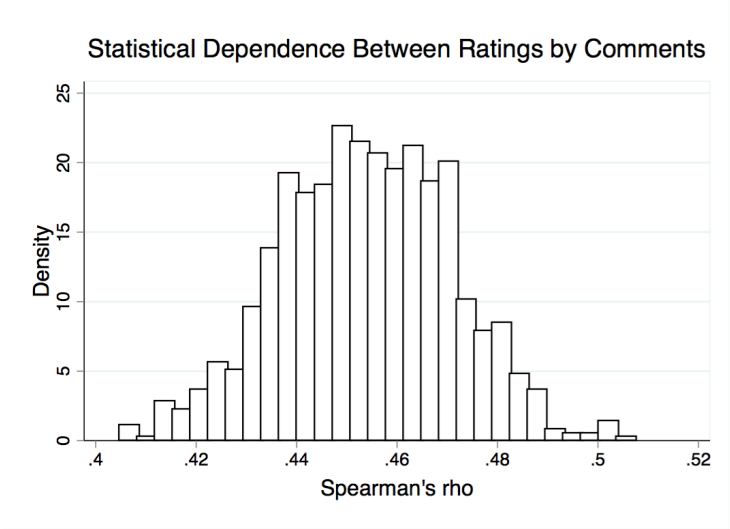


Figure A8: statistical dependence between the rankings

B.2 Additional Information on our Identification Strategy

A key idea of the identification strategy is the assumption that the terrorist attacks had an effect of online public discussion on refugees and related topics, but did not have an effect on the topics we use for comparison,, i.e., LGBT and feminism. We can compare the number of searches on the different topics after the terrorist attacks. Figure A9 shows the search interest of the terms Transgender, Feminism, LGBT and related topics relative to the highest point on the chart for Germany between 1 June 2017 and 31 August, 2016. The interest could be read as follows: a value of 100 is the peak of popularity for the term; a value of 50 means that the term is half as popular; likewise, a score of 0 means the term was less than 1% as popular as the peak. The popularity of these terms was not affected by the terrorist attacks (grey dashed lines). This can also be compared to the data in Figure 4.

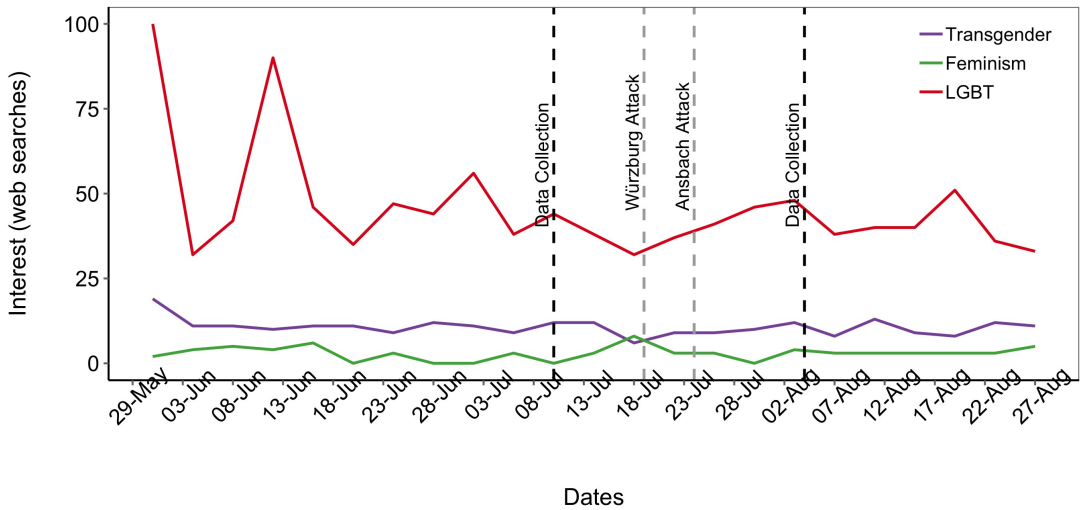


Figure A9: Relative interest (web searches) in Germany after the terrorist attacks. Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. The dashed lines represent the events of interest: dates of data collection and attacks.

B.3 Statistical Models

Our first statistical model is displayed in column 1 in Table 5 and is estimated only for the mixed condition using the comments on refugees:

$$Y_{ijk} = \beta_0 + \beta_1 AfterAttacks_{ijk} + u_i + v_j + \epsilon_{ijk} \tag{1}$$

where $u_i \sim N(0, \sigma_u)$ and $v_j \sim N(0, \sigma_j)$. The model estimates the rating Y_{ijk} for comment k by participant i on picture j. It is worth noting again that u_i (the random effect on the participant level) and v_j (the random effect on the picture level) are crossed random effects, that is pictures are neither fully nested in participants nor are participants fully nested in pictures. It means precisely that participants appear in more than one picture. This applies to all the following models except for the quantile regression.

The second model (Model 2 in Table 5) is specified as follows:

$$Y_{ijk} = \beta_0 + \beta_1 AfterAttacks_{ijk} + \beta_2 Topic_{ijk} + \beta_2(AfterAttacks_{ijk} \times Topic_{ijk}) + u_i + v_j + \epsilon_{ijk} \tag{2}$$

It introduces an interaction term between topic (Refugees or Gender Rights) and the terrorist attack.

Model 3 in Table 5 estimates the effects of the terrorist attacks in each experimental condition on comments about refugees:

$$Y_{ijk} = \beta_0 + \beta_1 AfterAttacks_{ijk} + \beta_2 Condition_{ijk} + \beta_3(AfterAttacks_{ijk} \times Condition_{ijk}) + u_i + v_j + \epsilon_{ijk} \tag{3}$$

Finally, the full model is displayed in column 4 of Table 5. It introduces three-way interactions between the attack, the topic, and the treatments and is estimated for the full data set:

$$\begin{aligned} Y_{ij} = \beta_0 + \beta_1 AfterAttacks_{ijk} + \beta_2 Topic_{ijk} + \beta_3 Condition_{ijk} + \beta_4(AfterAttacks_{ijk} \times Condition_{ijk}) \\ + \beta_5(AfterAttacks_{ijk} \times Condition_{ijk} \times Topic_{ijk}) + u_j + v_j + \epsilon_{ijk} \end{aligned} \tag{4}$$

The quantile regression models reported in subsection 6.3 have the standard form:

$$Y_i = \beta_0 + \beta_1(AfterAttacks_i \times Conditions_i) + u_i \tag{5}$$

The quantile regression estimator for quantile q minimizes the objective function:

$$Q(\beta_q) = \sum_{i:y_i \geqslant x_i\beta}^{N} q|y_i - x_i^{'}\beta_q| - \sum_{i:y_i \geqslant x_i\beta}^{N} (1-q)|y_i - x_i^{'}\beta_q| \tag{6}$$

We calculated quantile regression estimates for percentiles 0.10 to 0.95 in increments of 0.05.

B.4 Additional Results for the Quantile Regression

Figure A10 shows the coefficient of the effect of the terrorist attacks for the different topics and treatments for the quantiles 0.10 to 0.95 of the distribution of the hate score, as well as the density curves[27] before and after the terrorist attacks. The top left plot in Figure A10 shows the effect of the terror attack in the mixed treatment in Refugees. The terrorist attacks increased the average hate speech for all quantiles of the distribution (the mean increase is 0.56 points), but the effect is stronger in the most "hateful" quantiles of the distribution.

[27] A Gaussian kernel was used to estimate the density curves.

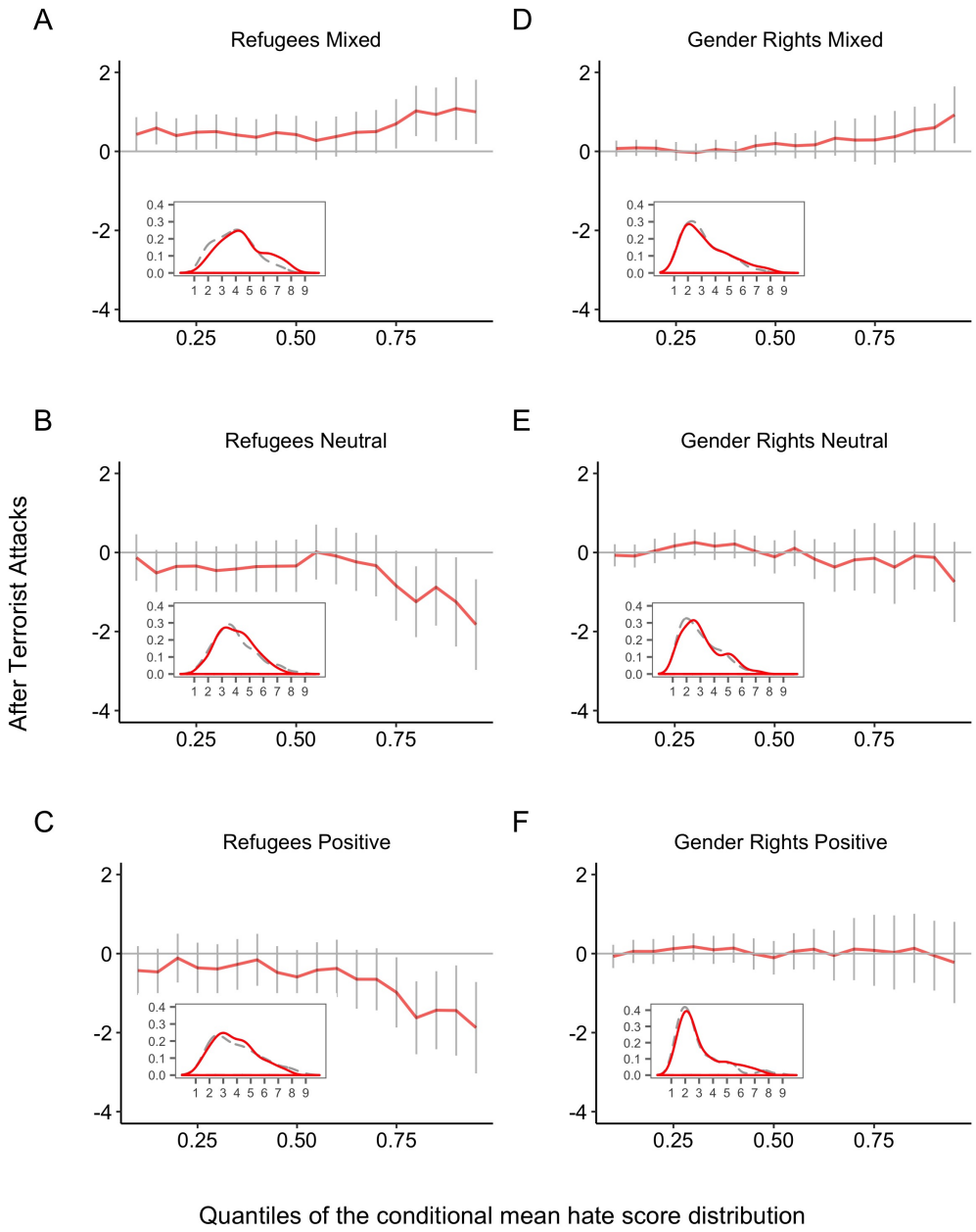


Figure A10: The plots depict the estimated model coefficients of the effect on terrorist attacks for quantiles 0.10 to 0.95 for each combination of topic and experimental condition. The grey vertical lines represent the confidence interval of the quantile regression coefficients for the effect of terrorist attacks. The plots also show the density distribution of the hate score before (dashed line) and after (solid line) the terrorist attacks. The left column shows the plots for Refugees. From top to bottom: mixed treatment (A), neutral treatment (B), and positive treatment (C). The column on the right shows the results for Gender Rights, from top to bottom: mixed treatment (D), neutral treatment (E), and positive treatment (F).

Table A2: Estimated Regression Coefficients for Percentiles 10 to 95 of the Distribution of the Hate Score of the Treatment Effect of the Terrorist Attacks in each Experimental Condition in Comments on Refugees. Confidence Intervals at 95% Level.

Quantile	Intercept	Mixed Condition Coef (up , lb)	Neutral Condition Coef (up , lb)	Positive Condition Coef (up , lb)
0.10	1.571	0.429(−0.009 , 0.866)	−0.129 (−0.716 , 0.459)	−0.429 (−1.047 , 0.189)
0.15	1.536	0.589(0.176 , 1.002)	−0.518 (−1.101 , 0.065)	−0.464 (−1.053 , 0.124)
0.20	2.200	0.400(−0.037 , 0.837)	−0.352 (−0.962 , 0.257)	−0.114 (−0.732 , 0.503)
0.25	2.229	0.486(0.042 , 0.929)	−0.343 (−0.972 , 0.286)	−0.361 (−0.998 , 0.277)
0.30	2.500	0.500(0.063 , 0.937)	−0.458 (−1.074 , 0.157)	−0.389 (−1.014 , 0.237)
0.35	2.833	0.417(−0.029 , 0.863)	−0.417 (−1.047 , 0.213)	−0.274 (−0.915 , 0.368)
0.40	3.143	0.357(−0.102 , 0.816)	−0.357 (−1.004 , 0.290)	−0.157 (−0.815 , 0.50)
0.45	3.190	0.476(0.012 , 0.940)	−0.348 (−1.005 , 0.309)	−0.476 (−1.142 , 0.190)
0.50	3.484	0.425(−0.050 , 0.899)	−0.341 (−1.011 , 0.330)	−0.591 (−1.271 , 0.089)
0.55	3.850	0.275(−0.217 , 0.767)	0.011 (−0.684 , 0.706)	−0.418 (−1.123 , 0.287)
0.60	3.875	0.375(−0.132 , 0.882)	−0.089 (−0.805 , 0.626)	−0.375 (−1.102 , 0.352)
0.65	3.893	0.482(−0.038 , 1.002)	−0.238 (−0.973 , 0.498)	−0.649 (−1.394 , 0.096)
0.70	4.000	0.500(−0.047 , 1.047)	−0.333 (−1.110 , 0.443)	−0.650 (−1.435 , 0.135)
0.75	4.179	0.696(0.073 , 1.320)	−0.839 (−1.725 , 0.046)	−0.982 (−1.870 , −0.094)
0.80	4.119	1.024(0.387 , 1.661)	−1.246 (−2.142 , −0.350)	−1.624 (−2.547 , −0.701)
0.85	4.467	0.933(0.251 , 1.615)	−0.878 (−1.850 , 0.095)	−1.433 (−2.421 , −0.446)
0.90	4.667	1.083(0.289 , 1.877)	−1.250 (−2.380 , −0.120)	−1.440 (−2.582 , −0.299)
0.95	5.500	1.000(0.188 , 1.812)	−1.825 (−2.972 , −0.678)	−1.875 (−3.031 ,−0.719)

Table A3: Estimated Regression Coefficients for Percentiles 10 to 95 of the Distribution of the Hate Score of the Treatment Effect of the Terrorist Attacks in each Experimental Condition in Comments on Gender Rights. Confidence Intervals at 95% Level

Quantile	Intercept	Mixed Condition Coef (up , lb)	Neutral Condition Coef (up , lb)	Positive Condition Coef (up , lb)
0.10	1.429	0.071 (−0.130 , 0.273)	−0.071 (−0.352 , 0.209)	−0.071 (−0.366 , 0.223)
0.15	1.536	0.089 (−0.115 , 0.294)	−0.089 (−0.381 , 0.202)	0.055 (−0.238 , 0.348)
0.20	1.698	0.079 (−0.138 , 0.297)	0.040 (−0.270 , 0.349)	0.054 (−0.262 , 0.370)
0.25	2.000	0.00 (−0.237 , 0.237)	0.167 (−0.165 , 0.498)	0.125 (−0.208 , 0.458)
0.30	2.175	−0.032 (−0.263 , 0.200)	0.254 (−0.077 , 0.585)	0.175 (−0.162 , 0.511)
0.35	2.238	0.048 (−0.198 , 0.294)	0.161 (−0.189 , 0.510)	0.095 (−0.261 , 0.451)
0.40	2.429	0.00 (−0.258 , 0.258)	0.214 (−0.151 , 0.580)	0.139 (−0.232 , 0.510)
0.45	2.429	0.143 (−0.132 , 0.418)	0.038 (−0.353 , 0.428)	−0.009 (−0.405 , 0.386)
0.50	2.600	0.200 (−0.096 , 0.496)	−0.111 (−0.531 , 0.309)	−0.105 (−0.530 , 0.320)
0.55	2.714	0.143 (−0.176 , 0.462)	0.107 (−0.346 , 0.560)	0.057 (−0.402 , 0.516)
0.60	3.000	0.167 (−0.189 , 0.523)	−0.167 (−0.672 , 0.339)	0.111 (−0.401 , 0.623)
0.65	3.000	0.333 (− 0.109 , 0.776)	− 0.369 (−0.995 , 0.257)	−0.048 (−0.681 , 0.586)
0.70	3.429	0.286 (−0.261 , 0.833)	−0.186 (−0.962 , 0.591)	0.114 (−0.672 , 0.900)
0.75	3.822	0.289 (−0.335 , 0.913)	−0.146 (−1.034 , 0.742)	0.082 (−0.815 , 0.980)
0.80	4.057	0.371 (−0.280 , 1.023)	−0.371 (−1.298 , 0.555)	0.029 (−0.910 , 0.967)
0.85	4.267	0.533 (−0.065 , 1.131)	−0.089 (−0.939 , 0.761)	0.133 (−0.740 , 1.007)
0.90	4.800	0.600 (−0.011 , 1.211)	−0.124 (−0.991 , 0.743)	−0.052 (−0.935 , 0.830)
0.95	4.950	0.925 (0.207 , 1.643)	−0.744 (−1.759 , 0.270)	−0.225 (−1.253 , 0.803)