

This is a post-print version of the article:

Kaakinen, M., Oksanen, A. & Räsänen, P. (2017). Did the Risk of Exposure to Online Hate Increase After the November 2015 Paris Attacks? A Group Relations Approach. *Computers in Human Behavior* 2017, vol. 78, pp. 90-97. <https://doi.org/10.1016/j.chb.2017.09.022>

Did the Exposure to Online Hate Increase After the November 2015 Paris Attacks? A Group Relations Approach**Abstract**

This study analyzed the impact of the November 2015 Paris attacks on online hate. On the basis of social identity based theories of group relations, we hypothesized that exposure to online hate will increase in social climate of fear, uncertainty, and polarization. We expected that the increase of hate will be evident in the case of online hate associated to ethnicity or nationality, religion, political views, or terrorism, but not specifically other hate-associated categories. Societal level determinants of the temporal changes in online hate exposure have not been tested before. Our study utilized two cross-sectional, demographically balanced datasets to analyze the change in online hate exposure among Finnish young people aged 15 to 30. The first sample was collected in May–June 2013 and the second one in December 2015, only 1 month after the November 2015 Paris attacks. The results supported the hypotheses indicating that the quantity and quality of online hostilities are affected by the wider societal conditions. We suggest that more evidence of societal level determinants of online hostility is needed in order to understand online hate exposure rates at different times.

Keywords: online hate, uncertainty–identity theory, social categories, social identity, groups, terrorism, immigration

Did the Exposure to Online Hate Increase After November 2015 Paris Attacks? A Group Relations Approach

Online hate (i.e., cyberhate, online hate speech, and other hate material) is a global phenomenon and may take many forms and target others based on their religion, race, ethnicity, gender, sexual orientation, national origin, or some other group-defining characteristic (Banks, 2010; Hawdon, Oksanen, & Räsänen, 2016; Perry & Olsson, 2009). Notably, online hate is not an exception to the rules of interaction in the online setting but rather rooted in mainstream experience, and exposure to online hate has varied from 31 percent to 67 percent in different samples across countries (Costello, Hawdon, Ratliff, & Grantham, 2016b; Hawdon et al., 2016; Oksanen, Hawdon, Holkeri, Näsi, & Räsänen, 2014).

The widely present hateful and xenophobic content online has raised concern in average online users but also in policymaking on a national and international level (Council of Europe, 2015; European Commission Against Racism and Intolerance, 2016; Gagliardone, Gal, Alves, & Martinez, 2015). Hostile online behavior bears hurtful consequences to its victims (Keipi et al., 2017; Näsi et al., 2015; Tynes, 2006; Ybarra, Mitchell, Wolak, & Finkelhorn, 2006), but it can also be considered a threat to societal inclusiveness and a potential motivator for hateful acts offline (Awan & Zempi, 2016; Douglas, 2007; Waldron, 2012). Especially in the times of social crises, such as terrorist attacks, online hate becomes an example of the increasing acts of violence and abuse faced by ethnic and religious minorities (Awan & Zempi, 2016).

When tackling online hate phenomena, we need wider empirical information on the prerequisites of online hostility. Earlier research has identified several correlates of violent online behavior on the levels of individual characteristics (e.g. low self-control, and high impulsivity, psychopathy or internalizing symptoms) and social interaction (e.g. anonymity,

low social control and group norms) (for review see e.g. Peterson & Densley, 2017). These correlates can explain why certain individuals and interactional contexts make hostile online behavior more probable. However, we lack empirical research on how wider societal (or macro) level phenomena can motivate changes in the quantity and quality of online hate over time and how this change is manifested in the viewership of such content.

Witnessing tragic and unexpected societal events may explain why manifestations of anger and hate take new forms online. Previous studies have shown that online discussions escalate after dramatic events such as rampage shootings (Lindgren, 2011). These types of attacks may also act as trigger events and direct contents of online hate. Williams and Burnap (2016) have recently demonstrated how particularly racial and religious cyberhate in Twitter escalated after a murder by Islamic extremists in the United Kingdom. However, earlier studies have stressed relatively short time periods and specific discussion topics on certain social media platforms. Thus, there is a need for research-based knowledge about the dynamics of online hate during longer periods of time and the wider viewership-centered point of view. Only this would allow us to assess how frequent is the experience of being exposed to online hate among social media users and whether the probability of exposure changes over time. This study is also the first one to approach temporal change in online hate from the perspective of group relations.

In this paper, we analyze how social conditions marked by fear, polarization, and uncertainty are manifested in online hate exposure after the terrorist attacks in Paris in 2015. On November 13, 130 people in Paris were killed in attacks by the terrorist organization ISIS, and the assault caused major societal reactions throughout Europe. The atmosphere in Europe was already insecure at the time, as several strikes by international terrorist organizations had occurred around the world that year (Haugerud, 2016). One devastating example in the European context was the attack on the satirical magazine *Charlie Hebdo* in January, which

led to the death of 12 people. These attacks also motivated antagonistic reactions toward immigrants, and concerns were raised that refugees were potential terrorists despite the fact that many of them were escaping the terror caused by ISIS in the Middle East (Nail, 2016). Immigration was already a matter of growing societal debate throughout Europe due to the so-called “immigration crisis” caused by the conflicts in Syria and Iraq, forcing over 1.2 million people to seek out asylum in Europe, and the figures of incoming first-time asylum seekers peaked during September–November 2016 (EuroStat, 2016).

Online Hate and Group relations

Our intergroup stance to online hate is based on the rich tradition of social psychological research explaining how prejudices are grounded on intergroup behavior (Allport, 1954; Brown, 2010; Tajfel, 1970). The starting point is grounded on previous empirical studies showing that online hate is typically targeted toward different social groups (Banks, 2010; Gagliardone et al., 2015; Hawdon et al., 2016; Perry & Olsson, 2009). Since the early days of domestic Internet, there have been both formal and informal hate groups disseminating hateful speech or ideology online. They have a wide variety of targets and ideological views, ranging from terrorist organizations to gangs of various types (Gerstenfeld, 2013, pp. 130–131).

Currently, different affordances of social media make it possible for people to group up with likeminded individuals without spatial restrictions and then disseminate their thoughts. Thus, social media can provide a social context of opinion congruence and empowerment in which people are more willing to express thoughts and ideologies that might be rejected elsewhere (Chun & Lee, 2017; Lee & Chun, 2015). This makes social media a particularly suitable platform for disseminating hateful or “fringe” opinions and ideologies (Barkun, 2017). In addition, the socio-technological environment of online interaction that

enhances the group identifications and the intragroup processes of online groups can legitimate and amplify extreme attitudes (Douglas, 2007; McGarty, Lala, & Douglas, 2011; Postmes, Spears, Sakhel, & De Groot, 2001; Spears, Postmes, Lea, & Wolbert, 2002). It is perhaps not a surprise that hate is often disseminated through those channels that make group formation and engagement very easy and that facilitate the clash between different ideological views (Erjavec & Kovačič, 2012; Hutchens, Cicchirillo, & Hmielowski, 2015).

Our theoretical framework of group relations is based on work done on social identity theory (SIT; Tajfel & Turner, 1979) and self-categorization theory (SCT; Turner, 1985; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). First, SIT suggests that individual identity is based on social categorization and comparison between different categories (Tajfel & Turner, 1979). People conceive themselves as members of certain groups (the in-groups) and as non-members in others (the out-groups) that strive to maintain positive social identity by favorable comparison between those groups. This search for self-enhancement leads to so-called intergroup bias in which the in-group is favored over the out-groups. The activation of intergroup bias is dependent on the level of individuals' identification with the in-group, the relevance of the comparison between the groups in a given social situation, and the relevance of the out-group as a reference point (Tajfel & Turner, 1979).

Second, according to SCT (Turner, 1985; Turner et al., 1987), an integral part of the social identity approach, identifying oneself as a representative of a certain category also leads to depersonalization (i.e. a tendency to conceive the self in terms of group identity instead of personal identity). As a consequence, one strongly identifies with the stereotypical conception of an in-group member and with the group attributes and norms (Brown, 2010; Marqués, Abrams, Paez, & Hogg, 2001; Turner, 1985; Turner et al., 1987).

In social reality, favoring one's in-group over perceived out-groups ranges from "mere categorical exaggerations" to extreme forms of out-group hostility (Billing, 2002,

178). Thus, a proper understanding of societal circumstances can help us to explain why and when biased perceptions of out-group members are likely to escalate. When societal conditions threaten the satisfaction of basic human needs, such as the need for security and control over one's own life, those out-groups perceived as responsible for the unsatisfactory circumstances are often targeted by increasing hostilities (Staub & Bar-Tal, 2003). Indeed, there is plenty of historical evidence showing how different group conflicts arise in times of fear, economic hardship, social and political segregation, and perceived in-group threats (see, e.g., Baumeister, 1997; Staub, 1989; Staub & Bar-Tal, 2003).

According to the terror management theory (TMT) (Greenberg, Pyszczynski, & Solomon, 1986), anxiety caused by the awareness of one's own vulnerability and death functions as a motivator to increase intergroup bias. In other words, people will seek to buffer the terror of mortality salience by more strongly identifying with worldviews shared within one's in-group. In addition, people are more likely to discriminate against out-groups that threaten, or do not validate, their anxiety-buffer (the in-group cultural worldview) (Greenberg, Solomon, Pyszczynski, & Lyon 1990). TMT has gained support from several studies reporting that mortality salience is related to out-group discrimination and support for extreme and violent attitudes (Burke, Martens, & Faucher, 2010; Das, Bushman, Bezemer, Kerkhof, & Vermeulen, 2009; Greenberg et al., 1990).

The theory of uncertainty-identity (UIT) (Hogg, 2007, 2014; Hogg et al., 2013) predicts that, at times of social uncertainty, individuals perceive the safety or manageability of their everyday life as endangered and, thus, tend to categorize the social reality according to more rigidly and exclusionary defined groups for overcoming the experienced uncertainty. As a consequence, the intra-group bias becomes inflated, leading to the adoption of more radical attitudes toward out-group members. A series of research has shown that, as a consequence of uncertainty, individuals tend to identify more with clearly distinctive groups

and approve more extreme ideologies and behaviors (Doosje, Lose man, & van den Bos, 2013; Esses, Stelian Medianu, & Lawson, 2013; Federico, Hunt, & Fisher, 2013; Grant & Hogg, 2012; Hogg, 2010; Hogg & Adelman, 2013).

In addition to mortality salience (TMT) and uncertainty (UIT), perceived threats from out-groups on physical well-being and power (realistic threat) or values and beliefs (symbolic threat) of the in-group tend to motivate out-group discrimination (Stephan, Ybarra, & Rios Morrison, 2009; Stephan & Renfro, 2002). In addition to realistic and symbolic threats, the intergroup threat theory (ITT) has been complemented with elements of group esteem threats, being personally threatened in intergroup interactions and having negative expectations concerning the out-group, for example (Brown, 2010; Koomen, W. & Van der Pligt, 2016; Stephan & Stephan, 2000). Indeed, perceived group threats have been reported to predict out-group discrimination in various studies (Kauff et al., 2015; Riek, Mania, & Gaertner, 2006; Wirtz, van der Pligt & Doosje, 2016).

Our theoretical framework suggests that societal events and conditions can shape the quantity and quality of online hate. Firstly, out-group hostilities will increase at times of fear, uncertainty and perceived threats. Secondly, people will witness more hostile messages in online space, as the social media provides an accessible forum for expressing and spreading hateful thoughts without the restrictions present in offline social networks and other media (Barkun, 2017). Finally, instead of all possible social categories, online hate is targeting categories perceived as relating to unsatisfactory conditions (Staub & Bar-Tal, 2003). The tendency of societal crises to escalate hateful reactions toward certain social categories in social media has already been demonstrated in the studies of online hate as well (see Awan & Zempi, 2016; Williams & Burnap, 2016).

Immigration Crisis and Social Uncertainty in Finland

Finland is a Nordic country with 5.5 million inhabitants and a relatively low ethnic diversity (United Nations, 2013). Finland, like other Nordic countries (Sweden, Norway, Denmark, and Iceland), has been characterized by a strong welfare tradition that includes free education and healthcare and extensive income distribution (Esping-Andersen, 1999).

Finland, like the rest of the world, was hit by the economic recession in 2007–2008, and the economy has been flat ever since. Economic problems have also concerned average citizens who have faced record figures of payment default (Oksanen, Aaltonen, & Rantala, 2015).

The so-called “immigration crisis” in Europe, caused by the conflicts in Syria and Iraq, started during the time when people already felt uncertain about the future due to the stagnating economy. During 2015, Finland received over five times as many asylum seekers as in the previous year, and the total amount of immigrants rose over 32,000. The peak of the crisis was during the autumn period, and different immigration centers were opened throughout the country. Clearly, the biggest group of immigrants came from Iraq (20,485), followed by Afghanistan (5,214), Somalia (1,981), and Syria (877).

The theme of immigration was a matter of heated debate in Finnish society (Silvennoinen, 2016). One example was the Rajat Kiinni (close the borders) Facebook group, which organized anti-immigration events starting from September 2015. Many of these anti-immigration events often targeted Islam as well. Along with anti-immigration events, the year 2015 in Finland witnessed a 50 percent increase in hate crimes and the rise of right-wing extremism and street patrol groups such as Soldiers of Odin that were organized via Facebook (Finnish Ministry of the Interior, 2016a). Accused of sexual harassment and rape cases, immigrants were also widely discussed in both traditional media and social media (Silvennoinen, 2016). Also, scandalous “fake media” online journals, such as *MV-lehti*, started to gain massive popularity.

The economic recession has also been visible in the Finnish societal immigration discourse (Keskinen, 2016). Traditionally, the Finnish national identity has been based on the Nordic egalitarian welfare model, but this policy has faced increasing pressure at the times of recession and growing demands for economic austerity in Finland. In some public opinions immigrants have been seen as a threat to the welfare state and some have requested more nationalistic and exclusive welfare policies that would deny the welfare benefits from the immigrants.

In social media, Finns were divided between those opposing immigration and those taking a more liberal stance. This is the result of a longer continuum in which the changing mediascape and new media technology have allowed for the formation of anti-immigration online communities (e.g., Hommaforum) and their social and political mobilization (Horsti, 2015). The escalated division between groups with anti-immigration stances and more tolerant ideologies has created an ongoing debate that is salient in Finnish online forums but also in mainstream media. The mainstream media was also often criticized for letting the anti-immigration movement set the agenda for public discussion and, of course, for staying silent about immigration-related topics by the anti-immigration movement (Horsti, 2015).

The year 2015 can be regarded as the year that tinged social confrontation and culminated in the Paris tragedy of November 13. The victims and the proximal consequences of the assault were located in Paris and France, but the media-based contact extended its psychosocial impact throughout Europe and the rest of the world. This phenomenon of “secondhand terrorism” (see Comer & Kendal, 2007) is manifested in extended uncertainty (i.e., a social climate that is determined by continuous insecurity and fear of future attacks based on possibilities instead of probabilities). This effect of extended uncertainty can be even stronger in countries with infrequent terrorist assaults such as Finland (Comer, Bry, Poznanski, & Golik, 2016). The social climate of fear is perhaps mirrored in the fact that the

fear of becoming a victim of violent or sexual assault among Finnish increased in 2015, a trend that had been decreasing in recent years (Finnish Ministry of the Interior, 2016b; Danielsson & Kääriäinen, 2016). This stressor has emerged despite crime occurrence not having changed during the given years.

Here, we treat the case of Finland as a context for a natural experiment analyzing the temporal change in online hate in times of social insecurity and uncertainty. As described before, the social climate in Finland, modified by economic recession, the peaking numbers of asylum seekers, and ideological and political divides, was already strained before the tragic events in Paris and its psychosocial impact. We believe that the increased salience of perceived safety and other threats along with social uncertainty is manifested in social identification and group discrimination processes as predicted by social psychological theories of group relations, TMT (Greenberg et al., 1986), UIT (Hogg, 2007, 2014) and ITT (Stephan et al., 2009).

Research Hypotheses

In this study, we analyze the temporal change in online hate phenomenon by utilizing two datasets collected in 2013 and 2015. Our presumption is that the quantity and quality of online hate is being shaped by the societal conditions. To test this presumption, we have formulated a total of three research hypotheses based on our theoretical framework:

First, we hypothesize that as a consequence of unsafety and uncertainty causing events in 2015, people will witness more hostility expressed in online spaces; thus, our survey respondents will report more exposure to hateful material online in 2015 when compared to 2013 (H1). Second, we hypothesize that people will report more exposure to hate associated with categories relating to social uncertainty and perceived in-group threats (ethnicity or nationality, religion, political views, and terrorism; H2). The third hypothesis is

that no similar effect will be found in terms of less relevant categories (gender, sexual orientation, disability, and appearance; H3).

The hypotheses are based on previous studies indicating the association between unsatisfactory societal conditions and intergroup hostilities (Hogg, 2007; Greenberg, Pyszczynski, & Solomon, 1986; Staub & Bar-Tal, 2003; Stephan, Ybarra, & Rios Morrison, 2009) and triggering societal events and hateful messaging in social media (Williams & Burnap, 2016).

Method

Participants

The first dataset was collected in May–June 2013 from Finland. The participants ($n = 555$) were aged 15 to 30 ($Mean = 22.59$, $SD = 4.21$), and half of them (50.0%) were female. The respondents were recruited from a panel of volunteer respondents administrated by Survey Sample International (SSI) and the sample was stratified to mirror the Finnish population aged 15 to 30 on a number of different socio-demographic factors, including age, gender, and living region.

The second dataset was collected in the beginning of December 10–15, 2015 approximately 1 month after the Paris attacks. Those respondents who were aged 16 to 30 were selected for this study ($n=192$, mean age = 23.13, $SD = 3.78$, 57.73% female). The data collection and respondent recruitment was administrated by the TNS Finland, and the sample was stratified to mirror the Finnish population in terms of age, gender, and living region. The used quotas allowed small differences to official population statistics.

Measures

Dependent variables. *Online hate exposure* was measured by asking participants whether they had seen (in the last 3 months) online material that threatened or degraded individual or social groups (“no” answers were coded as 0 and “yes” answers were coded as

1). Those participants who had seen online hate content were then presented with a follow-up question concerning the characteristics that the hateful material had related to. Options included ethnicity or nationality, religious conviction, political views, sexual orientation, gender, disability, appearance, and terrorism. This measure of online hate and its subtypes has been utilized in earlier research (Costello, Hawdon, & Ratliff, 2016a; Costello et al., 2016b; Hawdon et al., 2016; Oksanen et al., 2014). The exposure to these hate content types was measured with dummy variables (0 indicating no exposure and 1 indicating at least one exposure experience).

Treatment variable. Our treatment variable indicated whether the observation was made after the Paris attacks. The value of this variable was 0 for those respondents who had answered the survey before the attacks (the dataset collected in 2013) and 1 for respondents included in the dataset collected at the end of 2015.

Covariates. Our covariates (i.e, age, gender, education, living arrangements, and the activity of Internet use) were chosen on the basis of earlier studies concerning the associations between online hate exposure and sociodemographic characteristics and online activity (Costello et al., 2016b; Hawdon et al., 2016; Räsänen et al., 2016). Education was measured by including two dummy variables indicating whether the participant had completed secondary level education or higher level education (the reference category being basic level education). The variable measuring living arrangements indicated whether the participant was living with his or her parents. The activity of Internet use was measured with a dummy variable indicating whether the participant used the Internet more than once a day.

Analytic Strategy

Making inferences about the effect of certain interventions or events (i.e., treatment) on the basis of observational data is not unproblematic. The problems are often conceptualized as selection bias (see, e.g., Rickles, 2016), as the treated group (exposed to an

intervention or event) and the group it is being compared to (control group) may differ in terms of pre-treatment characteristics. This means that the observed outcomes for these groups cannot be attributed to the treatment, as they may as well be caused by other differences between them. Thus, causal relationships are often being studied by randomized experiments in which the comparability of treated and control groups is ensured by randomly placing individuals in those groups (Stuart, 2010). However, the randomized experiment design is not always accessible in the field of social sciences, and this is perhaps most prominent when examining events such as terrorist attacks. In these cases, there are analytical strategies for observational studies to replicate some characteristics of randomized experiment design (Stuart, 2010).

To account for possible selection bias, due to sociodemographic characteristics and online use activity, we used an analytical combination of the statistical technique of nearest-neighbor matching (see Rubin, 1973; Stuart, 2010) and logistic regression analysis in estimating the effect of the Paris attacks on hateful online content. With nearest-neighbor matching, we created two comparative sets of participants from the 2013 dataset (control group prior to the assault) and the 2015 dataset (the treated group). In the process, every respondent in the treatment group was given a closest matching counterpart from the control group. The selection of control group respondents was based on the propensity score, which is a measure of distance between the observations based on the likelihood of belonging to the treatment group. The likelihood is based on a logistic model using the selected covariates as predictors and the membership in the treated group as an outcome variable.

In our nearest-neighbor matching process, we used age, gender, education, living arrangements, and the activity of Internet use as covariates. As can be seen from Table 1, our covariate balance improved significantly as a consequence of the matching procedure. In the case of unmatched datasets, there are statistically significant differences in distributions of

gender, education, and online activity. In the 2015 data, there are more females, highly educated individuals, and active online users than there were in the data collected in 2013. The participants in the 2015 data, on average, are also older, and there are more of those who have completed a secondary degree education and are living with their parents, even though these estimates are not quite significant (with p-values of 0.071, 0.061, and 0.081, respectively). After the matching process, all differences in covariate distribution are statistically insignificant. In the matching process, six respondents from the treated group (originally 194) were discarded due to missing observations relating to samples of 188 participants.

<<Table 1 about here>>

In step two, we first combined the matched samples in one aggregated dataset. Then we estimated the effect of the Paris attacks (the treatment variable) as a change in the probability of online hate exposure by running a logistic regression model with the covariates also utilized in the propensity score estimation. When representing the results of this predictive analysis, we report the average marginal effects, standard errors, z-values, and p-values. This two-step model combining the matching process and regression analysis is a recommended way to reduce selection bias, and it has been used on earlier studies as well (see Klement, 2016; Rubin & Thomas, 2000).

Results

There was a substantive increase of online hate exposure among young Finnish people between 2013 and 2015. Approximately 47 percent of respondents had encountered online hate during 2013, while the proportion was 74 percent at the end of 2015. Table 2 also shows changes in the contents of online hate. In 2013, the most common form of online hate

concerned ethnicity or nationality (59.9%), followed by sexual orientation (57%) and appearance (42%). When measured after the Paris attacks, the hate was most frequently related to ethnicity or nationality (81%), religious conviction (71%), terrorism (66%), and political views (47%). For predictive analysis of the probability of online hate exposure after the Paris attacks, see Table 3.

<<Table 2 about here>>

According to our logistic regression analysis, there was a significant increase in the probability of exposure to online hate in general as well as to hate relating to categories of terrorism, religious conviction, ethnicity or nationality, political views, and gender. The probability of exposure to online hate in general was 27 percent higher after the Paris attacks. In the case of different subtypes of online hate, the effect was strongest for the probability of exposure to terrorism-related hate, which increased 42 percent. The probability of exposure to online hate relating to religious conviction was 34 percent higher at the end of 2015, while the probability increased 28 percent for ethnicity- or nationality-related hate and 19 percent for hate concerning political views. There was also a significant increase of 8 percent in the probability of exposure to hate related to gender categories. The effect estimates for hate targeting social categories of disability, sexual orientation, and appearance were not statistically significant.

<<Table 3 about here>>

Discussion

In this article, we analyzed how the exposure to online hate had changed in quantity and quality as a consequence of public uncertainty after the Paris attacks preceded by prolonged economic recession and escalated immigration debate. The measurement of online hate was done over the time period of 2013 and 2015, just a month after the Paris attacks. Our theoretical presumption was that out-group hostilities will increase at times of fear, uncertainty and perceived threat (Hogg, 2007; Greenberg, Pyszczynski, & Solomon, 1986; Stephan, Ybarra, & Rios Morrison, 2009) and, therefore, people will encounter more hate content in online space. In addition, we presumed that both the quantity and quality of online hate will change as hostilities will be targeted towards social categories perceived as related to fear, uncertainty and threats (Awan & Zempi, 2016; Staub & Bar-Tal, 2003; Williams & Burnap, 2016).

Analysis confirmed that encountering hate content online was more frequent after the Paris attacks, as young people were 27 percent more likely to be exposed to online hate at the end of 2015 than they were in the 2013 sample. The finding is in line with earlier research showing that, as a consequence of tragic societal events, hateful messaging will increase in social media (Awan & Zempi, 2016; Williams & Burnap, 2016). When interpreted through our theoretical framework of social psychological theories of group relations (Greenberg et al., 1986; Hogg, 2007, 2014; Stephan & Stephan, 2000), this result shows how times of insecurity and uncertainty tend to inflate the intergroup bias of favoring in-group members over the out-group members. As a result, people are more willing to adopt critical and even extreme stances and behavior toward out-groups. In online spaces, this is manifested in increased probability to encounter hostilities expressed toward social groups. It is also likely that individuals with hateful opinions perceive increasing presence in online hate content as

empowering which, in turn, further lowers the threshold of expressing one's own hostile thoughts (Barkun, 2017; Chun & Lee, 2017).

In addition, we expected that people would witness more hostility associated with social categories relating to the Paris attacks, as well as the political and ethnic divides preceding them (e.g. ethnicity or nationality, religious conviction, political views, terrorism). We also expected that no similar effect would be found in the case of categories not related to the social climate of extended uncertainty (e.g. gender, sexual orientation, disability and appearance).

Our analysis confirmed that young people encountered more online hate in categories related to a social climate of fear and uncertainty, but there was no similar effect in the case of nonrelated categories. According to our theoretical framework, the activation of intergroup bias is dependent on the relevance of the intergroup comparison and the out-group as a reference point, in addition to individuals' identification with the in-group (Tajfel & Turner, 1979). Here, categories of ethnicity or nationality, religious conviction, political views and terrorism have become more accessible references for comparison and conceptualization of social reality. Increased hostilities toward individuals and social groups perceived to share some features with perpetrators of criminal public events has also been reported in earlier research (Awan & Zempi, 2016; Williams & Burnap, 2016), even though there is obviously no actual connection between them.

Limitations

In our analyses, the problems relating to causal interference using observational data were taken into account by using a two-step analytical strategy in which a nearest-neighbor matching, based on estimated propensity scores, was first performed to enhance the comparability between two datasets (or the treatment and control groups). Then a logistic regression was conducted with the matched dataset for estimating the change in online hate

exposure after the Paris attacks. Even though this analytical strategy is a recommended way of assessing causal effects with observational data (Rubin & Thomas, 2000), the hypothesized causal relationship between the Paris attacks and the increase in online hate cannot be unquestionably confirmed by using cross-sectional data.

Another possible limitation of the present study is the subjective and category theme-based measurement of the online hate. This approach has, however, also been utilized elsewhere (Costello et al., 2016a; Costello et al., 2016b; Hawdon et al., 2016; Oksanen et al., 2014). For example, as we assessed how the self-reported exposure to hateful material online had changed between the two chosen time points, the increase in hate content exposure could be, at least partly, explained by some individuals becoming more sensitive to recognizing hostile online content.

On the basis of the category theme-based measurement, in turn, we cannot tell which specific groups were targeted by the encountered hostilities. However, we decided to utilize the category theme-based approach instead of the specific group-based one in order to analyze which social divides had escalated in social media. As there is a multitude of possible grouping criteria for political views, for example, our focus here was not to examine which ideological groups were those targeted by online hate but to analyze whether the political views were used as a social discrimination criterion in the first place. Of course, there is a long and highly important tradition of social psychological research examining xenophobic attitudes and prejudiced thinking toward specific social groups (see, e.g., Brown, 2010) as well as studies analyzing how certain groups are being exposed to online hate (for hate expressed toward the Muslim population, see, e.g., Awan & Zempi, 2016; Williams & Burnap, 2016). We believe there is a need for research-based knowledge on online hate generated by both approaches.

Conclusions

The hateful online content is apparent in social media interactions (Costello et al., 2016a; Hawdon et al., 2016; Oksanen et al., 2014). As the online hate is damaging for both individual victims (Keipi et al., 2017; Näsi et al., 2015; Tynes, 2006; Ybarra, Mitchell, Wolak, & Finkelhorn, 2006) and societies as collectives (Awan & Zempi, 2016; Waldron, 2012), it has become a target of national and international policy making and interventions (Council of Europe, 2015; European Commission Against Racism and Intolerance, 2016; Gagliardone et al, 2015). However, more research based knowledge is needed on why hostile online behavior emerges and how it changes over time. Individual and interaction level explanations (see Peterson & Densley, 2017) can reveal why some individuals and social contexts are more prone to online hostility. However, these explanations do not cover the fact that hate content spreading in social media appear to mainly attack certain social categories and this selectivity is shaped by temporal societal conditions. Thus, future research and policy interventions tackling the consequences of hostile online behaviour should stress how societal conditions always contribute to present and future forms of online hate.

References

- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Awan, I., & Zempi, I. (2016). The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. *Aggression and Violent Behavior, 27*, 1–8.
doi:10.1016/j.avb.2016.02.001
- Banks, J. (2010). Regulating hate speech online. *International Review of Law, Computers & Technology, 24*(3), 233–239.
- Barkun, M. (2017). President Trump and the ‘Fringe’. *Terrorism and Political Violence, 29*, 437–443.
- Baumeister, R. F. (1997). *Evil: Inside human cruelty and violence*. New York, NY: WH Freeman.
- Billig, M. (2002). Henri Tajfel's ‘Cognitive aspects of prejudice’ and the psychology of bigotry. *British Journal of Social Psychology, 41*(2), 171–188.
- Brown, R. (2010). *Prejudice: Its social psychology*. Chichester: John Wiley & Sons.
- Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two Decades of Terror Management Theory: A Meta-Analysis of Mortality Salience Research. *Personality and Social Psychology Review, 14*(2), 155–195.
- Chun, J. W. & Lee M. J.(2017). When does individuals’ willingness to speak out increase on social media? Perceived social support and perceived power/control. *Computers in Human Behavior, 74*, 120–129.
- Comer, J. S., & Kendall, P. C. (2007). Terrorism: The psychological impact on youth. *Clinical Psychology: Science and Practice, 14*(3), 179–212. doi: 10.1111/j.1468-2850.2007.00078.x
- Comer, J. S., Bry, L. B., Poznanski, B., & Golik, A. M. (2016). Children’s mental health in the context of terrorist attacks, ongoing threats, and possibilities of future terrorism.

Current Psychiatry Reports, 18(9), 79. doi:10.1007/s11920-016-0722-1

Costello, M., Hawdon, J., & Ratliff, T. (2016a). Confronting online extremism: The effect of self-help, collective efficacy, and guardianship on being a target for hate speech.

Social Science Computer Review, 1–19. Advance online publication.

doi:10.1177/0894439316666272

Costello, M., Hawdon, J., Ratliff, T., & Grantham, T. (2016b). Who views online extremism?

Individual attributes leading to exposure. *Computers in Human Behavior*, 63, 311–320.

Council of Europe. (2015). No hate speech movement: Campaign for human rights online.

Retrieved from <http://www.nohatespeechmovement.org/>

Danielsson, P. & Kääriäinen, J. (2016). Suomalaiset väkivallan ja omaisuusrikosten kohteena

2015 – Kansallisen rikosuhritutkimuksen tuloksia. Katsauksia 13/2016. Helsinki:

Helsingin yliopiston kriminologian ja oikeuspolitiikan instituutti.

Das, E. Bushman, B. J., Bezemer, M. D., Kerkhof, P., & Vermeulen, I. E. (2009). How

terrorism news reports increase prejudice against outgroups: A terror management account. *Journal of Experimental Social Psychology*, 45, 453–459.

Doosje, B., Loseman, A., & van den Bos, K. (2013). Determinants of radicalization of

Islamic youth in the Netherlands: Personal uncertainty, perceived injustice, and perceived group threat. *Journal of Social Issues*, 69(3), 586–604. doi:

10.1111/josi.12030

Douglas, K. M. (2007). Psychology, discrimination, and hate groups online. In A. Joinson, K.

McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 155–164). New York, NY: Oxford University Press.

- Esses, V. M., Medianu, S., & Lawson, A. S. (2013). Uncertainty, threat, and the role of the media in promoting the dehumanization of immigrants and refugees. *Journal of Social Issues, 69*(3), 518–536. doi: 10.1111/josi.12027
- Esping-Andersen, G. (1999). *Social foundations of postindustrial economies*. Oxford University Press.
- Erjavec, K., & Kovačič, M. P. (2012). You don't understand, this is a new war! Analysis of hate speech in news web sites' comments. *Mass Communication and Society, 15*(6), 899–920. doi: 10.1080/15205436.2011.619679
- European Commission Against Racism and Intolerance. (2016). ECRI general policy recommendation no. 15 on combating hate speech. Strasbourg, France: Council of Europe. Retrieved from www.coe.int/t/dghl/monitoring/ecri/activities/GPR/EN/Recommendation_N15/REC-15-2016-015-ENG.pdf.
- Eurostat. (2016). Asylum quarterly report. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Asylum_quarterly_report
- Federico, C. M., Hunt, C. V., & Fisher, E. L. (2013). Uncertainty and status-based asymmetries in the distinction between the “good” us and the “bad” them: Evidence that group status strengthens the relationship between the need for cognitive closure and extremity in intergroup differentiation. *Journal of Social Issues, 69*(3), 473–494.
- Finnish Ministry of the Interior (2016a). Väkivaltaisen ekstremismin tilannekatsaus 1/2016. Sisäministeriön julkaisu 23/2016. Helsinki: Sisäministeriö.
- Finnish Ministry of the Interior (2016b). Poliisibarometri 2016. Kansalaisten käsitykset poliisin toiminnasta ja sisäisen turvallisuuden tilasta. Sisäministeriön julkaisu 27/2016. Helsinki: Sisäministeriö.

- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). Countering online hate speech, (UNESCO series of Internet freedom). Retrieved from <http://unesdoc.unesco.org/images/0023/002332/233231e.pdf>
- Gerstenfeld, P. B. (2013). *Hate Crimes: Causes, Controls, and Controversies*. London: Sage.
- Grant, F., & Hogg, M. A. (2012). Self-uncertainty, social identity prominence and group identification. *Journal of Experimental Social Psychology, 48*(2), 538–542.
- Greenberg, J., Pyszczynski, X., & Solomon, S. (1986). The causes and consequences of a need for self-esteem: A terror management theory. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 189-212). New York: Springer-Verlag.
- Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A., Veeder, M., & Kirkland, S. (1990). Evidence for Terror Management Theory II: The Effects of Mortality Salience on Reactions to Those Who Threaten or Bolster the Cultural Worldview. *Journal of personality and Social Psychology, 58*(2), 308–318.
- Hawdon, J., Oksanen, A., & Räsänen, P. (2016). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior. early online*, [doi:10.1080/01639625.2016.1196985](https://doi.org/10.1080/01639625.2016.1196985)
- Hogg, M. A. (2007). Uncertainty-identity theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 69–126). San Diego, CA: Academic Press. [doi:10.1016/S0065-2601\(06\)39002-8](https://doi.org/10.1016/S0065-2601(06)39002-8)
- Hogg, M. A., Meehan, C., & Farquharson, J. (2010). The solace of radicalism: Self-uncertainty and group identification in the face of threat. *Journal of Experimental Social Psychology, 46*(6), 1061–1066.
- Hogg, M. A., & Adelman, J. (2013). Uncertainty–identity theory: Extreme groups, radical behavior, and authoritarian leadership. *Journal of Social Issues, 69*(3), 436–454.

- Hogg, M. A., Kruglanski, A., & van den Bos, K. (2013). Uncertainty and the roots of extremism. *Journal of Social Issues, 69*(3), 407–418.
- Hogg, M. A. (2014). From uncertainty to extremism social categorization and identity processes. *Current Directions in Psychological Science, 23*(5), 338–342.
- Horsti, K. (2015). Techno-cultural opportunities: The anti-immigration movement in the Finnish mediascape. *Patterns of Prejudice, 49*(4), 343–366.
- Hutchens, M. J., Cicchirillo, V. J., & Hmielowski, J. D. (2015). How could you think that?!?!: Understanding intentions to engage in political flaming. *New Media & Society, 17*(8), 1201–1219.
- Kauff, M., Asbrock, F., Issmer, C., Thörner, S., & Wagner, U. (2015). When immigrant groups “misbehave”: The influence of perceived deviant behavior on increased threat and discriminatory intentions and the moderating role of right-wing authoritarianism. *European Journal of Social Psychology 45*, 641–652.
- Keipi, T., Näsi, M. J., Oksanen, A. & Räsänen, P. (2017). *Online Hate and Harmful Content: Cross-National Perspectives*. London: Routledge.
- Keskinen, S. (2016). From welfare nationalism to welfare chauvinism: Economic rhetoric, the welfare state and changing asylum policies in Finland. *Critical Social Policy, 36*(3), 352–370.
- Klement, C. (2016). Outlaw biker affiliations and criminal involvement. *European Journal of Criminology, 13*(4), 453–472.
- Koomen, W. & Van der Pligt, J. (2016). *The psychology of radicalization and terrorism*. New York: Routledge.
- Lee, M. J., & Chun, J. W. (2016). Reading others’ comments and public opinion poll results on social media: Social judgment and spiral of empowerment. *Computers in Human Behavior, 65*, 479–487.

- Lindgren, S. (2011). YouTube gunmen? Mapping participatory media discourse on school shooting videos. *Media, Culture & Society*, 33(1), 123–136.
- Marqués, J. M., Abrams, D., Paez, D., & Hogg, M. A. (2001). Social categorization, social identification, and rejection of deviant group members. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell handbook of social psychology: Group Processes* (pp. 400–424). Oxford, UK: Blackwell.
- McGarty, C., Lala, G., & Douglas, K. (2011). Opinion-based groups: (Racist) talk and (collective) action on the Internet. In Z. Birchmeier, B. Deitz-Uhler, & G. Stasser (Eds.), *Strategic uses of social technology: An interactive perspective of social psychology* (pp. 145–171). Cambridge, UK: Cambridge University Press.
- Nail, T. (2016). A tale of two crises: Migration and terrorism after the Paris attacks. *Studies in Ethnicity and Nationalism*, 16(1), 158–167. doi: 10.1111/sena.12168
- Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., & Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Information Technology & People*, 28(3), 607–628. doi: 10.1108/ITP-09-2014-0198.
- Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. *Sociological Studies of Children & Youth*, 18, 253–273. doi:10.1108/S1537-466120140000018021
- Oksanen, A., Aaltonen, M., & Rantala, K. (2015). Social determinants of debt problems in a Nordic welfare state: A Finnish register-based study. *Journal of Consumer Policy*, 38(3), 229–246. doi: 10.1007/s10603-015-9294-4.
- Perry, B., & Olsson, P. (2009). Cyberhate: The globalization of hate. *Information & Communications Technology Law*, 18(2), 185–199.
- Peterson, J. & Densley, J. (2017). Cyber violence: What do we know and where do we go from here? *Aggression and Violent Behavior*, 34, 193–200.

- Postmes, T., Spears, R., Sakhel, K., & De Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin, 27*, 1242–1254
- Rickles, J. (2016). A review of propensity score analysis: Fundamentals and developments. *Journal of Educational and Behavioral Statistics, 41*(1), 109–114.
- Riek, B. M., Mania, E. W., & Gaertner, S. (2006). Intergroup threat and outgroup attitudes: a meta-analytic review. *Personality and Social Psychology Review, 10*(4), 336–353.
- Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics, 29*(1), 159–184.
- Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95*(450), 573–585.
- Räsänen, P., Hawdon, J., Holkeri, E., Näsi, M., Keipi, T., & Oksanen, A. (2016). Targets of online hate: Examining determinants of victimization among young Finnish Facebook users. *Violence & Victims, 31*(4), 708–726.
- Silvennoinen, O. (2016). But—where do these people come from? The (re) emergence of radical nationalism in Finland. Retrieved from <http://www.sicherheitspolitik-blog.de/2016/04/04/but-where-do-these-people-come-from-the-reemergence-of-radical-nationalism-in-finland/>
- Spears, R., Postmes, T., Lea, M., & Wolbert, A. (2002). When are net effects gross products? The power of influence and the influence of power in computer-mediated communication. *Journal of Social Issues, 58*, 91–107.
- Staub, E. (1989). *The roots of evil: The psychological and cultural origins of genocide and other forms of group violence*. New York, NY: Cambridge University Press.
- Staub, E. & Bar-Tal, D. (2003). *Genocide, Mass Killing and Intractable Conflict: Roots,*

- Evolution, Prevention, and Reconciliation. In: D. O. Sears, L. Huddy, and R. Jervis (Eds.), *Oxford Handbook of Political Psychology* (pp. 710–751). New York: Oxford University Press.
- Stephan, W. G., & Renfro, C. L. (2002). The role of threats in intergroup relations. In D. Mackie and E. R. Smith (Eds.), *From prejudice to intergroup emotions* (pp. 191-208). New York: Psychology Press.
- Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing Prejudice and Discrimination* (pp. 23-45). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stephan, W. G., Ybarra, O., & Rios Morrison, K. (2009). Intergroup Threat Theory. In T. Nelson (Ed.), *Handbook of Prejudice* (pp. 43-59). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21.
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96–102.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks Cole.
- Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (pp. 77–122). Greenwich, CT: JAI.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Blackwell.

- Tynes, B. (2006). Children, adolescents, and the culture of online hate. In N. E. Dowd, D. G. Singer & R. F. Wilson (Eds.), *Handbook of Children, Culture, and Violence* (pp. 267–289). Thousand Oaks, CA: Sage.
- United Nations. (2013). Trends in international migrant stock: The 2013 revision. Retrieved from <http://esa.un.org/unmigration/TIMSA2013/migrantstocks2013.htm?mtotals>
- Waldron, J. (2012). *The harm in the hate speech*. Cambridge, MA: Harvard University Press.
- Williams, M. L., & Burnap, P. (2016). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, *56*(2), 211–238. doi: 10.1093/bjc/azv059
- Wirtz, C., van der Pligt, J., & Doosje, B. (2016). Negative Attitudes Toward Muslims in The Netherlands: The Role of Symbolic Threat, Stereotypes, and Moral Emotions. *Peace and Conflict: Journal of Peace Psychology*, *22*(1), 75–83.
- Ybarra, M. L., Mitchell, K. J., Wolak, J., & Finkelhor, D. (2006). Examining characteristics and associated distress related to Internet harassment: findings from the Second Youth Internet Safety Survey. *Pediatrics*, *118*(4), e1169-e1177.

Table 1. The balance of selected covariates between 2013 and 2015 samples with mean values, standard deviations and p-values of the t-test

Un-matched samples

	2013		2015		t-test
	mean	sd	mean	sd	p-value
Age	22.59	4.21	23.18	3.76	0.071
Gender (0 = male, 1 = female)	0.50	0.50	0.59	0.49	0.041
Education					
Secondary (0 = no, 1 = yes)	0.50	0.50	0.57	0.50	0.061
Higher (0 = no, 1 = yes)	0.17	0.37	0.28	0.45	0.003
Living with parents (0 = no, 1 = yes)	0.32	0.47	0.25	0.43	0.081
Active online user (0 = no, 1 = yes)	0.73	0.44	0.92	0.28	0.000
n		555		188	

Matched samples

	2013		2015		t-test
	mean	sd	mean	sd	p-value
Age	23.44	3.80	23.18	3.76	0.504
Gender (0 = male, 1 = female)	0.61	0.49	0.59	0.49	0.600
Education					
Second level (0 = no, 1 = yes)	0.57	0.50	0.57	0.50	1.000
Higher education (0 = no, 1 = yes)	0.29	0.45	0.28	0.45	0.819
Living with parents (0 = no, 1 = yes)	0.24	0.43	0.25	0.43	0.811
Active online user (0 = no, 1 = yes)	0.90	0.30	0.92	0.28	0.720
n		188		188	

Table 2. Exposure to online hate in general and by different hate categories (frequencies and percentages)

	2013		2015	
	n	%	n	%
Seen online hate	88	46.8	139	73.9
<i>Categories related to uncertainty</i>				
Ethnicity or nationality	58	56.9	112	80.6
Religious conviction	33	37.5	98	70.5
Political views	29	33.0	65	46.8
Terrorism	12	13.6	92	66.2
<i>Categories not related to uncertainty</i>				
Sexual orientation	58	56.9	45	32.4
Appearance	37	42.0	35	25.2
Gender	22	25.5	37	26.6
Disability	14	15.9	16	11.5
n	188		188	

Table 3. Average partial effects (APEs) of Paris attacks on online hate in general and on different hate categories

	APE	SE	z-score	p-value
Seen online hate	0.268	0.047	5.649	0.000
<i>Categories related to uncertainty</i>				
Ethnicity or nationality	0.282	0.048	5.866	0.000
Religious conviction	0.341	0.044	7.683	0.000
Political views	0.191	0.043	4.404	0.000
Terrorism	0.422	0.039	10.738	0.000
<i>Categories not related to uncertainty</i>				
Sexual orientation	-0.073	0.046	-1.607	0.108
Appearance	-0.013	0.040	-0.341	0.733
Gender	0.075	0.037	2.042	0.041
Disability	0.008	0.028	0.280	0.780