Kaakinen Markus, Räsänen Pekka, Näsi Matti, Minkkinen Jaana, Keipi Teo, Oksanen Atte

Social Capital and Online Hate Production: A Four-Country Survey

Hateful, threatening or degrading content has become a common part of today's online interactions. However, little is known about the people who produce such content. This study analyzes online hate content production and its associations with cognitive indicators of social capital in both offline and online social networks. The data are derived from American, Finnish, German and British Internet users aged 15–30 (N=3,565). Measures included questions concerning online hate, social capital and contextual control variables. The results indicate that hate content production is rare overall, despite its high visibility, and is related to social capital in two key ways. First, respondents with high social capital in offline social networks were less likely to produce hate content, and second, high social capital in online networks was associated with a higher probability of production.

Keywords: online interaction; social media; social networks; online hate; social capital

Introduction

The Internet, and social media especially, have become central to information dissemination and personal expression in both a positive and negative sense in Western societies (Davidson & Martellozzo, 2013). Although a relatively small proportion of all messaging in social media can be considered hateful (Williams and Burnap 2015), hateful, threatening or degrading content is widespread and evident in online interaction (Brown 2009; Oksanen et al. 2014; Waldron 2012). The area of harmful content in focus here, namely online hate (i.e. cyberhate), involves the dissemination of racist or xenophobic content (Banks 2011; Council of Europe 2017a, 2017b; Waldron 2012) and acts that threaten or degrade individuals or social groups (Hawdon et al. 2014; Oksanen et al. 2014; Keipi et al., 2017). Mitigating the costs of this material is central to a safe online experience; however, while the prevention and surveillance of aggression and tensions in social media network communication are regarded as a significant policy aim (e.g. Council of Europe 2017a, 2017b), this remains a challenge for authorities and public policy making in terms of daily usage (Wall and Williams 2013; Williams et al. 2013). What is needed is a deeper look into user characteristics associated with content production in order to bolster preventative measures.

Various groups and individuals have been active online for some time and have utilized information and communications technology (ICT) to express and spread ideology and recruit new members, while attracting like-minded individuals and

targeting young people especially (Lee and Leets 2002; Wong et al. 2015). In the United States, white supremacist groups have been active on the Internet since the 1980s (Gerstenfeld et al. 2003; Levin 2002). These early adopters of spreading hate-based ideology online have been followed by scores of others. Right-wing populist movements and other proponents of extremism in Europe have also become more visible in the online context during the 2000s (Caiani and Parenti 2013; Lucassen and Lubbers 2012). In addition, the role of various Islamist terrorist groups, such as ISIS, has recently become increasingly active in leveraging social media services to promote their ideology (see e.g. Brown 2009; Klausen 2015).

However, not all threatening and degrading rhetoric online is disseminated solely by organized groups. Individuals acting on their own initiative also take part in the phenomenon (see e.g. Erjavec and Kovačič 2012). Furthermore, in addition to hate sites themselves, online hate has become widely present in mainstream social media. While Ybarra and colleagues reported in their study that only a small minority of American children and adolescents have visited so called hate sites (3.5% in 2008) (Ybarra et al. 2011), various recent studies have reported the presence of online hate in social media platforms (e.g. Costello et al., 2016; Erjavec and Kovačič 2012; Hawdon, Oksanen, Räsänen, 2016; Näsi et al., 2016; Oksanen et al. 2014; Sureka et al. 2010; Williams and Burnap 2015). Thus, exposure has spread past the early days when hate material was witnessed by those specifically seeking it out. Furthermore, cross-national comparisons show that online hate targets a variety of different topics, ranging from political and societal issues (e.g. ethnicity, political views, religious beliefs) to subjective issues (e.g. sexual orientation, physical appearance and disability) (Keipi, Näsi, Oksanen, & Räsänen, 2017). As such, this represents a broadening of the hate spectrum that is easily accessible online.

Existing hate content on the Internet has been studied extensively from a variety of methodological perspectives, including qualitative studies (Brown 2009; Gerstenfeld et al. 2003; Glaser et al. 2002), sophisticated web crawling techniques (Chau and Xu 2007; Wong et al. 2015), and survey studies focusing on victimization experiences and their implications (Näsi et al. 2016; Keipi et al. 2017). However, thus far few studies have actually examined hate content producers, particularly from a cross-national perspective relying on comprehensive survey data.

Toward filling this research gap, in this paper we examine online hate content production and its associations with cognitive indicators of social capital in both the offline and online social networks. In other words, we are focused on the extent to which perceived quality social relations to different social networks relates to the probability of online hate content production. Our focus will be on late adolescents and young adults from four countries, namely the US, UK, Germany and Finland.

Social relations online and social capital

The Internet provides various possibilities for forming social networks based on mutual interests independent of locality (Wellman and Gulia 1999). Among young people, social media and social networking sites (SNS) have become increasingly popular contexts for building identities (Lehdonvirta and Räsänen 2011), expressional platforms (Keipi and Oksanen 2014) and online communities (De Souza and Preece 2004; Kim and Kim 2017; Kingod et al. 2017). In addition, SNS offer affordances for person-to-person and person-to-network communication that help people to maintain contact with various offline social networks such as old high school friends or former colleagues, for example (Boyd and Ellison 2008; Hampton 2016; Hampton et al. 2011). This shift

underpins changes in wider community structures, namely from localities (e.g. neighborhoods) to personal, or egocentric, social networks of non-local social ties; though, it is important to note that this trend began well before the proliferation of social media (Boase and Wellman 2005; Wellman 1979; Hampton and Wellman 2003).

In the field of criminology and in the social sciences overall, the value of social networks to both individuals and collectives is typically conceptualized as social capital (De Silva et al. 2005, pp. 619; see also Oxoby 2009; Rosenfeld et al. 2001; Salmi and Kivivuori 2006; Sampson and Raudenbush 1997). Though varying approaches to social capital exist (see Bourdieu 1984; Coleman 1988; Portes 2000; Putnam 1993, 2000), the theoretical starting point of this study is Robert Putnam's (1993, 2000) definition of social capital as social networks, norms or reciprocity and trust that facilitate the social organization and civic virtue of societies. Furthermore, social capital has an internal value for community members in forms of social support or information access, for example. It also carried external or public value in the sense that social capital generated within communities also facilitates the large-scale organization of societies outside of themselves (Putnam and Goss 2002).

According to Wellman and his colleagues (2001), Putnam's approach to social capital can be divided into components of participatory capital and network capital. Participatory capital refers to the element of political and civic involvement that fosters collective action. Network capital, in turn, refers to individuals' ties to varying social networks (Hampton and Wellman 2003; Wellman et al. 2001). In this study, we define social capital, in line with the network capital, as a resource that consists of individuals' supportive ties to multitude of social networks providing individuals various psychological, social, and material benefits (e.g. emotional and material aid) (Hampton and Wellman 2003; Wellman et al. 2001).

In earlier research, social capital has been approached from the structural or cognitive perspective, where the structural component stresses network attributes (e.g. density of social network) and behavioral characteristics (e.g. social or political participation), while the cognitive component emphasizes individuals' subjective assessment of the quality of their social relations (e.g. trust of sense of belongingness) (De Silva et al. 2005; Harpman et al. 2002; Yip et al. 2007). Overall, social capital has been reported to associate with various positive outcomes on both the individual and collective level. Individuals with more social capital have, for example, higher self-rated health and life satisfaction (Han 2013; Kawachi et al. 1999; Elgar et al. 2011), whereas social capital on the collective level has been found to have a negative association with violence and criminality (Rosenfeld et al. 2001; Kennedy et al. 1998). As such, the concept of social capital has become an integral part of national and international public health policies and discourses in the new millennium (De Silva et al. 2005; Muntaner et al. 2000).

Returning to the social and expressional scene so common today, social networking sites have been found to assist in the creation and maintenance of social capital (Best and Krueger 2006; Bouchillon 2014; Ellison et al. 2007). SNS allows users to communicate with all of their personal networks while also utilizing the social capital embedded in them toward gaining support and resources from these social networks (Boase 2008; Ellison et al. 2014; Hampton and Wellman 2003; Lu and Hampton 2016). In addition, social networks formed and maintained online serve as a linkage between offline and online domains by offering platforms for bringing online social capital to the offline context and vice versa (Boase and Wellman 2005; Shen and Cage 2015; Rosen et al. 2011). In the online context, social capital embedded in online communities facilitates their internal functioning as network members with high social capital tend to

be the most active participants and information disseminators within the community, for example (Chang and Chuang 2011; Yen 2016).

Putnam and other communitarian theorists of social capital usually underline the public value of social capital as fostering large-scale social organization and civic engagement outside the networks as well (Julien 2014). Similarly, social capital in the online context is a resource that facilitates online civic engagement, such as free information sharing and collective problem solving, in online communities (Ibid.). This perspective to social capital, however, does not pay equal attention to elements of distinction, exclusion and conflict in and between social networks (Dally and Silver 2008; Julien 2014; Muntaner et al. 2000). Contrary to this, in the agonistic tradition (see e.g. Bourdieu 1984, 1986), social capital has been conceived as a resource for mutual recognition within social networks but also for conflict, distinction and exclusion between individuals and social groups (Julien 2014; Portes 2000). Indeed, some conflict features of online interaction within online communities, e.g. trolling or conflict behavior, are perhaps more understandable when viewed from the agonistic perspective (Julien 2014; McCosker 2014)

However, the communitarian view on social capital does not assume that social capital always functions for the public good (e.g. Putnam and Goss 2002). Some forms of social networks may indeed be beneficial for their members but destructive on the larger scale. For example, social capital can facilitate the social organization of crime (Putnam 1993), civil war parties (Pérez-Díaz 2002) or, in the extreme, genocides (McDoom 2014). Obviously, these forms of social capital do not function as a public good.

This study

In this study, our interest lies in whether social capital embedded in offline and online personal networks facilitates the large-scale social organization of online interaction, i.e. reduces anti-social online behavior. In our empirical analysis, we apply cognitive indicators of social capital, i.e. trust and sense of belongingness, to offline and online personal networks (a group of friends and to an online community). Here, we analyze the relation between online hate content production and social capital. Notably, this relationship remains unexplored despite the problem of online hate being widely discussed. We believe that this new perspective is important for a deeper understanding of hateful, threatening or degrading online communication. As such, we aim to fill the gap in the research with the following research questions:

RQ1: Does online hate content production associate with social capital embedded in offline and online personal networks?

RQ2: Do the detected associations indicate positive or negative relationships between online hate content production and social capital in offline and online networks?

We present two competing hypotheses concerning the direction of possible associations:

H1: Social capital embedded in offline and online networks will have public value in the sense that it will facilitate the social organization of online interaction (Putnam and

Goss 2002). Therefore, trust and a sense of belongingness to a group of friends and online community will have a negative association with conflict behavior, i.e. online hate content production.

H2: In online interaction, social capital embedded in offline and online networks will function as a resource in conflict. Given the distinction and exclusion between individuals and social groups (Julien 2014), trust and a sense of belongingness to a group of friends and an online community will have a positive association with conflict-based online activity, i.e. online hate content production.

Cross-national context

The respondents in our study come from the United States, Finland, United Kingdom and Germany. All four are technologically advanced countries with high general living standards. Of these four countries, the U.S is the most populated with more than 300 million inhabitants. The least populated nation, Finland, has a population of only 5.5 million people, with Germany (82 million) and the UK (63 million) falling in between. Income differences are smallest in Finland followed by Germany, the UK and then the U.S., which has the highest gross domestic product but also the largest income gap (Inequality Watch 2014). With only around 5 per cent immigration at the population level, Finland is also the most ethnically homogeneous of these nations, followed by Germany (12%), UK (13%) and then the U.S. (14%) (United Nations 2013).

In previous comparisons, Finland and the U.S. have been found to share relatively similar exposure rates to online harassment (Näsi et al. 2014). The U.S., however, differs from European countries when it comes to legislation on hate speech and hate content. In European countries, publishing threatening or abusive content directed against particular groups of people has been constrained through legislation, while in the US, legislation has favored the freedom of speech over hate content restrictions (Waldron 2012). Of European countries, Germany has particularly strict legislation on the matter which, for example, prohibits the denial or approval of the Holocaust (see Allen and Norris 2011; Brugger 2002; Hawdon, Oksanen and Räsänen 2016).

In international comparisons on generalized trust, for example, Finns are more likely to agree that people can be trusted compared to Germans, American or British on the scale of "Most people can be trusted or you can't be too careful" (OECD 2011). In the sixth round of the European Social Survey from 2012, Finns evaluated the item on average at a value of 6.6 compared to 5.5 and 5.2 for British and German respondents (on scale 1–10) (ESS 2012). Moreover, in an international comparison, American and British respondents showed more pro-social behavior than did Finnish and Germans, while all four nations perform around the average when it comes to measures of anti-social behavior (OECD 2011).

Methods

Data

To answer the addressed research questions, we use four comparable datasets drawn from youth and young adults (aged between 15 and 30). The data for Finland (n=555)

and the U.S. (n=1,033) were collected in the spring of 2013 while the data for Germany (n=978) and the UK (n=999) were collected in the spring of 2014. The questionnaires were identical for all countries, excluding language. Survey Sample International (SSI) administered the data collection. Email invitations were sent to a sample of SSI panel members recruited through random digit dialing, banner ads, and other permission-based techniques. The country samples were demographically balanced on gender, vocation and living situation. The survey items dealt with respondents' online activity and interactions as well as socio-demographic and socio-psychological characteristics.

Measures

Online hate content production was measured with a survey item in which respondents were asked whether they had produced online material that someone had interpreted as threatening or degrading with either a yes or no response.

    Social capital. Both offline and online social capital were assessed with cognitive indicators, i.e. as trust and a sense of belongingness. Trust and sense of belongingness have been used as a measure of cognitive aspects of social capital in earlier studies (Yip et al. 2007; Nyqvist et al. 2013) and validated in the Adapted Social Capital Assessment Tool (ASCAT) and a short version of the Adapted Social Capital Assessment Tool (SASCAT) (Harpham et al. 2002; De Silva et al. 2006). Trust was measured by survey items in which respondents were asked to rate whether they agreed with the following phrases: "good friends can be trusted" (offline trust) and "people only met on the Internet can be trusted" (online trust) on a 1–10 scale (1=you cannot be too careful, 10=can be trusted entirely). Sense of belongingness was measured by items in which respondents were asked to rate how closely they felt belonging in their circle of friends (offline sense of belongingness) and in an online community (online sense of belongingness) on a 1–5 scale (1=not closely at all, 5=very closely).

    Control variables: Gender, age, living alone and SNS activity were used as control variables. Living alone was measured with a survey item in which respondents were asked whether they lived alone, in a marriage or cohabitation with or without children, as a single parent, with parents, or according to some other arrangement (e.g. with flat mates). The answers "living alone" were coded as 1 (others as 0). SNS activity was measured with a survey item in which respondents were asked which social media technologies they had used during the past three months. The 21-item list included the most popular sites and services ranging from Facebook, Twitter and YouTube, to blogs, message and image boards, online game communities and communication tools (e.g. Skype, instant messengers, email). Those services that were rarely reported or only used in the national context (e.g. IRC-Galleria in Finland and MeinVZ in Germany) were left out of the analysis. A count variable that combined indicated services was then calculated (ranging between 1–19).

Statistical methods

Our data analysis is based on descriptive statistics and explanatory logistic regression models. We first give an overview of the phenomenon by focusing on differences and similarities among the four countries. We then continue our explanatory analyses with the aggregated data given the similarities among the countries. Explanatory analysis focuses on examining the associations between the individual-level independent variables and online hate production. When interpreting the results of our regression models, we report average marginal effects (AME) with standard errors (SE) rather than

somewhat more conventional odds-ratios. This is mainly because we are comparing the effects of individual variables across different models. AME coefficients provide us with reliable predictions from a model; in our case the probability of producing online hate content, while also taking into account other independent variables. We also chose to report average marginal effects, since the sizes of the effect will differ across the values of different independent variables in the models.

Results

Our descriptive results are given in Table 1. According to the results, online hate content production is a relatively rare phenomenon in all countries. While the US shows the highest proportion (with 4.1%) and Germany the lowest (with only 0.9%), the proportion of hate producers is three per cent for the entire sample. In this respect, there is no considerable variation in our samples. This can also be said when looking at the subjective independent variables. For example, country averages for offline trust are relatively close to one another, with Finland showing the highest mean of 7.8 and the UK showing the lowest with 7.2. The same is true in terms of offline belongingness averages (from 4.0 for Germany to 3.6 for the Finland). Furthermore, the levels of online trust appear to be relatively stable (with 3.7 for Finland and the UK and 3.3 for Germany). In terms of online belongingness, averages vary from 2.3 in Germany and the UK to 2.7 in Finland. In general, social capital is higher in the traditional offline context than in the online setting in all samples.

Descriptive analysis of socio-demographic control variables shows that our samples are well balanced in terms of gender, age, and SNS activity as well. Half of the respondents are female and the average age of respondents is 23 years. Respondents in Germany show a slightly lower average for SNS activities with 5.5 while the averages in other countries are approximately 6.2. When looking at the distributions of living arrangements, we observe that nearly 30 per cent of Finnish respondents live alone. The shares are clearly lower for other countries, varying from 12 per cent in the US sample to 20 per cent in Germany. These figures are in line with official statistics (Eurostat 2010; United States' Census Bureau 2015).

<< Table 1 about here>>

Given the similarities between the countries and the infrequency of reported hate content production, we continue our explanatory analysis by using the aggregated data. Results of our logistic regression analyses are presented in Table 2. The first model includes socio-demographic control variables and SNS activities. The second model adds subjective independent variables into the equation. We will examine both of the models separately, relying on AME coefficients. In the tables, the AMEs can be interpreted as decimal shares of how much the chances of producing hate content are expected increase or decrease for a unit change in the explanatory variables.

Model 1 shows that gender and living arrangement variables are significantly associated with the probabilities of online hate production. Other variables fail to be significant. Males are approximately 3 per cent more likely to produce hate content than females. On the other hand, those who live alone have a 2.5 per cent higher probability of producing such content when contrasted with other respondents.

Model 2 shows the findings from the adjusted model, which takes into account the differences associated with variables measuring social capital. The findings show

that living alone no longer significantly increases online hate production (probability declines from 2.5% to 1.5%). Gender, on the other hand, remains significant with males having a 2.2% higher probability of producing hate content. Offline trust decreases the probability of producing hate content (0.7% by a unit change in a 10-point scale), whereas online trust increases the probability (with 1.2% by a unit change in a 10-point scale) of doing so. Furthermore, online belongingness increases the probability of hate production by approximate 1.2 per cent by one unit in a five-point scale.

<< Table 2 about here>>

In general, male respondents reported significantly more production of online hate content than women. In addition, living alone increased the probability of hate content production in Model 1. Other control variables, namely age and SNS activity, did not have significant associations with hate production. What is notable is that offline and online social capital have different relations to online hate content production. Both online trust and sense of belongingness predicted a higher probability of hate content production. Offline social capital had an opposite, and less clear, association with online hate content production. As such, offline trust predicted a lower probability of hate content production. Finally, sense of belonging in a social group of friends was not significantly associated with online hate production.

Discussion

This article examined the possible association between online hate content production and social capital in offline and online personal networks. We stated two competing hypotheses concerning the direction of possible association. First (H1), we assumed that offline and online networks would be negatively associated with conflict behavior (Putnam and Goss 2002). Second (H2), we assumed that trust and a sense of belonging to a group of friends and an online community would have a positive association with conflict activity online, namely online hate content production (Julien 2014).

According to our descriptive results, even though hate content is widespread and common in social media, a relatively small group of people actually produces such material. This result is consistent with William and Burnap's (2015) finding that even in the case of increased hate messaging on Twitter after the Woolwich terrorist attack, only a small proportion of all messaging on the subject was actually hateful. There are, however, cross-national differences in the prevalence of online hate production, with the United States having a highest frequency (4.1%) and Germany (0.9%) the lowest. The distinctively low rate of Germany is perhaps unsurprising given country's particularly strict legislation on hate speech (see Allen and Norris 2011; Brugger 2002).

Our predictive analysis revealed that social capital embedded in offline and online social networks is associated with online hate content production. However, the connection is inverse; strong social capital in offline social network was negatively associated with online hate production. This association, though, was not unambiguous since only trust in friends was significantly associated with hate content production. The sense of belongingness to a group of friends did not have such a connection with the studied phenomenon. On the other hand, social capital embedded in an online social network actually increased the probability of hate content production. The more respondents trusted people only met online and the more they felt that belonging to an

online community, the more likely they were to produce hateful content in the online setting. Hence, it appears that our study supports both hypotheses concerning the direction of association between online hate content production and social capital. Here, social capital in the offline context acted, as presumed by the public value hypothesis (H1), as an integrative force and thus lowered the likelihood of conflict behavior online. On the contrary, in the case of social capital in the online context, our results support the conflict hypothesis (H2) in that those individuals with more online social capital are also more likely to act in a conflictive manner in online interaction.

Social capital embedded in personal networks online and offline appear to have differing implication for the wider social setting. According to our results, social capital embedded in online social networks does not have a public effect of reducing anti-social online behavior on average despite possibly having other external effects (for internal and external value see Putnam and Goss, 2002). Notably, offline social networks appear to have such an effect.

According to Putnam (Putnam 1993; Putnam & Goss, 2002), social capital is likely to enhance the overall cohesion, functionality and civic virtue of societies, but this is not always the case. Social capital within a certain community can also lead to bigotry or abuse towards other groups, as examples of organized crime (Putnam, 1993), civil wars (Pérez-Díaz, 2002) and even genocides (McDoom, 2014) point out. This is in line with Bourdieu's (Julien, 2014; Bourdieu 1984, 1986) agonistic view of social capital as a mutual source of approval and recognition but also as a resource in the struggles between different groups.

One possible explanation for the online social capital association with an increased likelihood of hostile online behavior is the nature of social media as a provider of easy access to both like-minded individuals and to a wide audience, including those with opposing opinions. Here, trust and a sense of belongingness to a certain online community might contribute to hostilities when views or ideologies are strongly approved within the like-minded community, but objected to elsewhere. This can result in extreme confrontation (see e.g. Erjavec and Kovačič, 2012). In addition, online community members with high social capital are, on average, most active in information sharing and communication within the community (Chang and Chuang 2011; Yen 2016). This tendency might increase the likelihood of ending up in conflict with other users in online interaction due to the normalization of validation in customized social contexts making differences less tolerable.

On the basis of this study, we cannot identify the motivation behind the produced hate content. Thus, it is not possible to indicate whether the producers act as members of organized online hate groups or according to a more individual-centered agenda. However, it appears that online and offline social networks matter when it comes to online hate.

Of our control variables, only gender was significantly associated with the production of online hate content in logit models, with men reporting a higher probability than women. This result is in line with earlier results on gender differences in aggressive behavior, though the difference has been found to be smaller for verbal aggression (Archer 2004). Furthermore, living alone first predicted a higher likelihood of producing hate content. The association, however, disappeared after adding the measures of social capital in the analysis. In other words, adolescents and young adults living alone but having equally strong offline and online social relationships were not more likely to produce online hate. Neither age nor the amount of social media activity was significantly associated with the studied phenomenon. This can be considered surprising since the quantity of Internet use has been found to predict exposure to online

hate (Oksanen et al. 2014) along with other online risks (Livingstone and Helsper 2010; Livingstone and Haddon 2009; Livingstone and Haddon 2008).

Limitations

Our study has some limitations. Firstly, we used cross-sectional data, which does not allow us to make causal attributions of the studied phenomena. Even though our theoretical framework presumes that social capital precedes the production of online hate content, we cannot present the results as cause and effect on the basis of our analysis. Instead, we focused on the associations between those two.

Secondly, though the infrequency of reported online hate content production might not be considered surprising (see e.g. Williams and Burnap 2015) this demands some theoretical and methodological reflection. It is possible that our results were affected by the so-called social undesirability bias (Tourangeau and Yan 2007), referring to the underreporting of socially sensitive matters by respondents not wanting to admit to socially reprehensible behavior. However, in the present study, social undesirability bias was taken into account by self-administering the survey and offering a private answering setting with no interviewer present, as respondents were able to answer the online survey from their own computer or mobile device. These utilized techniques have been found to reduce social undesirability bias (ibid.). Due to the small proportion of online hate production, we were not able to analyze hate content production separately for each country. This would have enabled us to examine and compare them as different sociocultural contexts for online hate. However, as shown in earlier research and indicated by the descriptive results of this study, those countries also share many features relevant in this context.

Finally, we also examined the effect of offline and online social capital by analyzing them separately. However, it should be noted that we are not suggesting that these are two distinct phenomena. As such, the possible associations and interactions between offline and online social capital, in terms of online hate, remains a relevant and new area of research for future study.

Conclusion

Hateful and aggressive online communication has been widely considered a social problem requiring adequate policy interventions (e.g. Council of Europe 2017a, 2017b). At the same time, online aggression remains a challenging phenomenon to manage for authorities and public policy makers (Wall and Williams 2013; Williams et al. 2013). According to our results, it appears that even though hateful content is apparent in social media, the production of online hate is a rare phenomenon and is related to social capital in inverse ways: strong social relations to online communities increase the likelihood of producing online hate content while traditional (offline) relations decrease the risk. In other words, enhanced inclusion to online social networks does not lead to increased integration in online interaction. However, as our results concerning offline social capital suggest, in addition to legislation and social media-driven measures and campaigns against online hate, authorities and policy makers should take into account the creation and maintenance of social capital in more traditional offline social networks as well when tackling hostile behavior in the online context.

References

Allen J and Norris G (2011) Is Genocide Different? Dealing with Hate Speech in a Post-Genocide Society. *Journal of international law and international relations 7*(1): 146–174.

Archer J (2004) Sex Differences in Aggression in Real-World Settings: A Meta-Analytic Review. *Review of General Psychology* 8(4): 291–322.

Banks J (2011) European Regulation of Cross-Border Hate Speech in Cyberspace: The Limits of Legislation. *European Journal of Crime, Criminal Law and Criminal Justice* 19(1): 1–13.

Best SJ and Krueger BS (2006) Online Interactions and Social Capital. Distinguishing Between New and Existing Ties. *Social Science Computer Review* 24(4): 395–410.

Boase J (2008) Personal Networks and the Personal Communication System. *Information Communication and Society 11*(4): 490–508.

Boase J and Wellman B (2005) Personal Relationships: On and Off the Internet. In: Perlman D and Vangelisti AL (eds.) *Handbook of Personal Relations*. Oxford: Blackwell, pp.709-723.

Bouchillon BC (2014) Social Ties and Generalized Trust, Online and in Person: Contact or Conflict—The Mediating Role of Bonding Social Capital in America. *Social Science Computer Review* 32(4): 506–523.

Bourdieu P (1984) *Distinction: A Social Critique of the Judgment of Taste.* Cambridge, MA: Harvard University Press.

Bourdieu P (1986) The Forms of Capital. In: Richardson JG (ed) *The Handbook of Theory and Research for Sociology of Education.* New York: Greenwood, pp.241–258.

Boyd DM and Ellison NB (2008) Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13(1): 210–230.

Brown C (2009) 'WWW.HATE.COM: White Supremacist Discourse on the Internet and the Construction of Whiteness Ideology.' *The Howard Journal of Communications* 20(2): 189–208.

Brugger W (2002) Ban On or Protection Of Hate Speech-Some Observations Based on German and American Law. *Tulane European & Civil Law Forum* 17(1): 1–21.

Caiani M and Parenti L (2013) *European and American extreme right groups and the Internet.* Farnham: Ashgate Publishing.

Chau M and Xu J (2007) Mining Communities and Their Relationships in Blogs: A Study of Hate Groups. *International Journal of Human-Computer Studies* 65: 57–70.

Chang HH and Chuang SS (2011) Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. *Information & Management 48*(1): 9–18.

Coleman JS (1988) Social Capital in the Creation of Human Capital. *American Journal of Sociology 94,* Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure: 95–120.

Costello M, Hawdon J, Ratliff T, and Grantham T (2016) Who views online extremism? Individual attributes leading to exposure. *Computers in Human Behavior 63*: 311–320.

Council of Europe (2017a) *No hate speech movement: Campaign for human rights online.* Available at: http://www.nohatespeechmovement.org/ (accessed 17 January 2017).

Council of Europe (2017b) European Commission against Racism and Intolerance (ECRI). Available at: http://www.coe.int/t/dghl/monitoring/ecri/default_en.asp (accessed 17 January 2017).

Dally M and Silver H (2008) Social exclusion and social capital: A comparison and critique. *Theory and Society* 37(6): 537–566.

Davidson J, Martelozzo E (2013) Exploring young people's use of social networking sites and digital media in the internet safety context. *Information, Communication & Society* 16: 1456–1476. doi: 10.1080/1369118X.2012.701655

De Silva MJ, Harpham T, Tuan T, Bartolini R, Penny M and Huttly S (2006) Psychometric and cognitive validation of a social capital measurement tool in Peru and Vietnam. *Social Science & Medicine* 62(4): 941–953.

De Silva MJ, McKenzie K, Harpham T and Huttly S (2005) Social capital and mental illness: a systematic review. *Journal of Epidemiology and Community Health* 59(8): 619–627.

De Souza CS and Preece J (2004) A framework for analyzing and understanding online communities. *Interacting with Computers* 16(3): 579–610.

Elgar F, Davis CD, Wohl MJ, Trites SJ, Zelenski JM and Martin MS (2011) Social capital, health and life satisfaction in 50 countries. *Health & Place* 17(5): 1044–1053.

Ellison NB, Gray R, Lampe C and Fiore A (2014) Social capital and resource requests on Facebook. *New Media & Society 2014* 16(7): 1104–1121.

Ellison NB, Steinfield C and Lampe C (2007) The Benefits of Facebook ''Friends':' Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication* 12(4): 1143–1168.

Erjavec K, Kovačič MP (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.

ESS (2012) *European Social Survey: Round 6* [Data file]. Available at: http://www.europeansocialsurvey.org/data/download.html?r=6 (accessed 21 November 2015).

Eurostat (2010) *Household structure in the EU*. 2010 Edition. Available at: http://ec.europa.eu/eurostat/documents/3888793/5848337/KS-RA-10-024-EN.PDF (accessed 21 November 2015).

Gerstenfeld PB, Grant DR and Chiang C (2003) Hate online: A content analysis of extremist internet sites. *Analysis of Social Issues and Public Policy* 3(1): 29–44.

Glaser J, Dixit J and Green DP (2002) Studying hate crime with the Internet: What makes racists advocate racial violence? *Journal of Social Issues* 58(1): 177–193.

Han S (2013) Compositional and contextual associations of social capital and self- rated health in Seoul, South Korea: A multilevel analysis of longitudinal evidence. *Social Science & Medicine, 80*: 113–120.

Hampton KN (2016) Persistent and Pervasive Community: New Communication Technologies and the Future of Community. *American Behavioral Scientist* 60(1): 101–124.

Hampton KN, Lee CJ, and Her EJ (2011) How new media affords network diversity: Direct and mediated access to social capital through participation in local social settings. New Media & Society 13(7): 1031–1049.

Hampton KN and Wellman B (2003) Neighboring in Netville: How the Internet Supports Community and Social Capital in a Wired Suburb. *City & Community, 2*(4), pp.277-311.

Harpham T, Grant E and Thomas E (2002) Measuring social capital within health surveys: Key issues. *Health Policy Plan* 17(1): 106–111.

Hawdon, J., Oksanen, A., & Räsänen, P. (2014). Victims of Online Groups: American Youth's Exposure to Online Hate Speech. In J. Hawdon, J. Ryan & M. Lucht (Eds.), *The Causes and Consequences of Group Violence: From Bullies to Terrorists* (pp. 165–182). Lanham, Maryland: Lexington Books – Rowman & Littlefield.

Hawdon J, Oksanen A, and Räsänen, P. (2016) Exposure to Online Hate in Four Nations: A Cross-National Consideration. *Deviant Behavior, 38(3), 254–266.* https://doi.org/10.1080/01639625.2016.1196985

Inequality Watch (2014) The rise of income inequality amongst rich countries[e-publication]. Retrieved from http://inequalitywatch.eu/spip.php?article58

Julien C (2014) Bourdieu, Social Capital and Online Interaction. *Sociology.* Epub ahead of print June 30 2014. DOI: 10.1177/0038038514535862
1–18. DOI: 10.1177/0038038514535862

Kawachi I, Kennedy BP and Glass R (1999) Social capital and self-rated health: a contextual analysis. *American Journal of Public Health* 89(8): 1187–1193.

Keipi, T., & Oksanen, A. (2014). Self-exploration, anonymity and risks in the online setting: Analysis of narratives by 14–18-year olds. Journal of Youth Studies, 17(8), 1097– 1113. https://doi.org/10.1080/13676261.2014.881988

Keipi, T., Näsi, M., Oksanen, A., & Räsänen, P. (2017). *Online Hate and Harmful Content: Cross-National Perspectives*. Abingdon & New York: Routledge. https://doi.org/10.4324/9781315628370

Kennedy BP, Kawachi I, Prothrow-Stith D, Lochner K and Gupta V (1998) Social capital, income inequality, and firearm violent crime. *Social Science and Medicine* 47(1): 7–17.

Kim MS and Kim HM (2017) The effect of online fan community attributes on the loyalty and cooperation of fan community members: The moderating role of connect hours. *Computers in Human Behavior,* 68, 232–243. https://doi.org/10.1016/j.chb.2016.11.031

Kingod N, Cleal B, Wahlberg A, and Husted GR (2017) Online Peer-to-Peer Communities in the Daily Lives of People With Chronic Illness: A Qualitative Systematic Review. *Qualitative Health Research* 27(1): 89–99.

Klausen J (2015) Tweeting the Jihad: Social media networks of Western foreign fighters in Syria and Iraq. *Studies in Conflict & Terrorism* 38(1): 1-22.

Lee E and Leets L (2002) Persuasive storytelling by hate groups online: Examining its effects on adolescents. *American Behavioral Scientist* 45(6): 927–957.

Levin B (2002) Cyberhate: A legal and historical analysis of extremists' use of computer networks in America. *American Behavioral Scientist* 45(6): 958–986.

Lehdonvirta V and Räsänen P (2011) How do young people identify with online and offline peer groups? A comparison between UK, Spain and Japan. *Journal of Youth Studies* 14(1): 91–108.

Livingstone S and Helsper EJ (2010) Balancing opportunities and risks in teenagers' use of the internet: the role of online skills and internet self-efficacy. *New Media & Society* 12(2): 309–329.

Livingstone S and Haddon L (2008) Risky Experiences for Children Online: Charting European Research on Children and the Internet. *Children and Society* 22(4): 314–23.

Livingstone S and Haddon L (2009) Introduction. In: Livingstone S and Haddon L (eds) *Kids Online. Opportunities and Risks for Kids*. Bristol, UK: The Policy Press, pp.1–15.

Lu W and Hampton KN (2016) Beyond the power of networks: Differentiating network structure from social media affordances for perceived social support. New Media & Society online first: 1–19. DOI: 10.1177/1461444815621514

Lucassen G and Lubbers M (2012) Who fears what? Explaining far-right-wing preference in Europe by distinguishing perceived cultural and economic ethnic threats. *Comparative Political Studies* 45(5): 547–574.

McCosker A (2014) Trolling as provocation: YouTube's agonistic publics. *Convergence: The International Journal of Research into New Media Technologies* 20(2): 201–217.

McDoom OS (2014) Antisocial capital: A profile of Rwandan genocide perpetrators' social networks. *Journal of Conflict Resolution* 58(5): 865–893.

Muntaner C, Lynch J and Smith GD (2000) Social capital and the third way in public health. *Critical Public Health* 10(2): 107–124.

Näsi, M., Räsänen, P., Oksanen, A., Hawdon, J., Keipi, T., & Holkeri, E. (2014). Association between online harassment and exposure to harmful online content: A cross-national comparison between the United States and Finland. Computers in Human Behavior, 41, 137– 145. https://doi.org/10.1016/j.chb.2014.09.019

Näsi, M., Räsänen, P., Kaakinen, M., Keipi, T., & Oksanen, A. (2017). Do routine activities help explain young adults' online harassment: A multi-nation study. Criminology & Criminal Justice, 17(4), 418– 432. https://doi.org/10.1177/1748895816679866

Nyqvist F, Forsman AK and Cattan M (2013) A comparison of older workers' and retired older people's social capital and sense of mastery. *Scandinavian Journal of Public Health* 41(8): 792–798.

OECD (2011) *Society at a Glance 2011: OECD Social Indicators*. Paris: OECD Publishing. DOI: http://dx.doi.org/10.1787/soc_glance-2011-en

Oksanen A, Hawdon J, Holkeri E, Näsi M and Räsänen P (2014) Exposure to Online Hate Among Young Social Media Users. *Sociological Studies of Children & Youth* 18: 253–273.

Oxoby R (2009) Understanding social inclusion, social cohesion, and social capital. *International Journal of Social Economics* 36(12), 1133–1152.

Pérez-Díaz V (2002) From civil war to civil society: Social Capital in Spain from the 1930s to the 1990s. In: Putnam R (ed) *Democracies in Flux: The Evolution of Social Capital in Contemporary Society*. New York, NY: Oxford University Press, pp.245–288.

Portes A (2000) The two meanings of meanings of social capital. *Sociological Forum* 15(1): 1–12.

Putnam R (1993) *Making Democracy Work: civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Putnam R (2000) *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.

Putnam R and Goss KA (2002) Introduction. In: Putnam R (ed) *Democracies in Flux: The Evolution of Social Capital in Contemporary Society*. New York, NY: Oxford University Press, pp.3–20.

Rosen D, Lafontaine PR and Hendrickson B (2011) CouchSurfing: Belonging and trust in a globally cooperative online social network. *New Media & Society* 13(6): 981–998.

Rosenfeld R, Messner SF and Baumer EP (2001) Social Capital and Homicide. *Social Forces* 80(1): 283–309.

Salmi V and Kivivuori J (2006) The Association between Social Capital and Juvenile Crime The Role of Individual and Structural Factors. *European Journal of Criminology* 3(2): 123–148.

Sampson RJ and Raudenbush SW (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* 277: 918–924.

Shen C and Cage C (2015) Exodus to the real world? Assessing the impact of offline meetups on community participation and social capital. *New Media & Society* 17(3): 394– 414.

Sureka A, Kumaraguru P, Goyal A, and Chhabra S (2010) Mining YouTube to discover extremist videos, users, and hidden communities. In: Cheng P-J, Kan M-Y, Lam W., and Nakov P (eds) *Information retrieval technology, volume 6458 of lecture notes in computer science*. Berlin, Germany: Springer, pp.13–24.

Tourangeau R and Yan T (2007) Sensitive questions in surveys. *Psychological Bulletin* 133(5): 859–883.

United States' Census Bureau (2015) *Young Adults Then and Now.* Available at: http://www.census.gov/censusexplorer/censusexplorer-youngadults.html (accessed 21 November 2015).

United Nations (UN, 2013) Trends in International Migrant Stock: The 2013 Revision. New York. Available at: http://esa.un.org/unmigration/TIMSA2013/migrantstocks2013.htm?mtotals (accessed 21 November 2015).

Waldron J (2012) *The harm in the hate speech*. Cambridge, MA: Harvard University Press.

Wall DS and Williams ML (2013) Policing cybercrime: networked and social media technologies and the challenges for policing. *Policing and Society* 23(4): 409–412.

Wellman B (1979) The Community Question: The Intimate Networks of East Yorkers. *American Journal of Sociology 84*(5): 1201-31.

Wellman B (2001) Physical place and cyber place: The rise of personalized networking. *International Journal of Urban and Regional Research 25*(2): 227-252.

Wellman B and Gulia M (1999) Virtual communities as communities: Net surfers Don't Ride Alone. In: Kollock P and Smith M (eds) *Communities and Cyberspace*. New York: Routledge, pp.167–194.

Williams ML, and Burnap P (2015). 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

Williams ML, Edwards A, Housley W, Burnap P, Rana O, Avis N, Morgan J and Sloan L (2013) 'Policing Cyber-Neighbourhoods: Tension Monitoring and Social Media Networks'. *Policing and Society* 23(4): 461–81.

Wong MA, Frank R and Allsup R (2015) The supremacy of online white supremacists – an analysis of online discussions by white supremacists. *Information & Communications Technology* 24(1): *41–73.*

Ybarra ML, Mitchell KJ and Korchmaros JD (2011) National Trends in Exposure to and Experiences of Violence on the Internet Among Children. *Pediatrics* 128(6): 1376–1386.

Yen C (2016) How to unite the power of the masses? Exploring collective stickiness intention in social network sites from the perspective of knowledge sharing. *Behaviour & Information Technology, 35*(2): 118–133.

Yip W, Subramanian SV, Mitchell AD, Lee DTS, Wang J and Kawachi I (2007) Does social capital enhance health and well-being? Evidence from rural China. *Social Science & Medicine* 64(1): 35–49.

Table 1. Descriptive statistics by country (variable means).

		Total	US	Finland	Germany	UK
	Range					
Has produced hate content	0=no, 1=yes	0.030	0.041	0.040	0.009	0.034
Independent variables						
Offline trust	1 - 10	7.43	7.38	7.76	7.55	7.17
Offline belongingness	1 - 5	3.78	3.79	3.57	3.98	3.69
Online trust	1 - 10	3.56	3.62	3.66	3.25	3.74
Online belongingness	1 - 5	2.42	2.50	2.74	2.32	2.26
Control variables						
Gender	0=male, 1=female	1.498	1.503	1.499	1.501	1.490
Age	15 - 30	23.07	23.09	22.59	23.20	23.18
Living alone	0=no, 1=yes	0.174	0.124	0.297	0.201	0.132
NSN activity	0 - 19	6.02	6.20	6.18	5.54	6.22
n		3565	1033	555	978	999

Table 2. Predicting online hate production in all four countries. Binary logit models, average marginal effects.

	Model 1			Model 2		
	Marginal effects	SE	p	Marginal effects	SE	p
Female	-0.027	0.006	0.000	-0.022	0.006	0.000
Age	-0.001	0.001	0.930	-0.001	0.001	0.935
Living alone	0.025	0.009	0.007	0.015	0.008	0.056
SNS activity	0.001	0.001	0.234	0.001	0.001	0.166
Offline trust				-0.007	0.002	0.000
Offline belongingness				-0.001	0.003	0.907
Online trust				0.005	0.001	0.001
Online belongingness				0.012	0.003	0.000
Pseudo R2	0.043			0.131		
n	3565			3519		