Backgrounds, Experiences and Responses to Online Hate Speech: An Ethnographic Multi-sited Analysis

Olga Jubany[1]

Serra Hunter Fellow, Director of European Social Research Unit, Department of Social Anthropology, Universitat de Barcelona, Montalegre, 6-8. 08001 Barcelona, Spain

Abstract. The complexity of our global economic, social and political climate has provided a platform for exacerbating tensions between heterogeneous groups, often expressed through increasing levels of hate speech, fostered by intolerance and racism. With the emergence of different social media platforms, such tensions are reflected, and even reinforced, by the large scope and impacts of online media. This is especially the case for young people, often defined as the front runners in the use of digital communication tools. Grounded on an ethnographic multi-sited study, this paper discusses how people experience and react to online hate speech, and argues how a dominating global pattern of normalisation of this phenomenon is expanding, particularly amongst young people.

Introduction

Inequalities across our contemporary heterogeneous societies are on the rise, powerfully affecting young people globally. At the same time, the emergence of diverse social media platforms has revolutionised social interaction patterns, where young people tend to be defined as the front runners in the use of digital communication tools. However, the current difficult economic, social and political climate has provided an ideal platform for exacerbating tensions between constructed groups, often expressed through hate speech, fostered by increasing levels of intolerance and racism. Whereas emerging online spaces provide many new possibilities for interaction and personal expression, these are also increasingly flooded by the problematic of online hate speech[2]. Whilst xenophobic and racist hate speech is not a new phenomenon, through the advanced use of internet and social media, new ways of dissemination and expansion have emerged. Young social media users are at risk of coming across hate messages on social media platforms, giving rise to a critical need to look deeper into young people's perceptions of and responses to this type of content.

The problematics begin with the difficulties in conceptualising "hate speech". Whilst this is a term widely used, a variety of different interpretations are applied, very much depending on the time and space contexts. In legal terms, there has long been a search for a precise definition, as the most accepted one has proved to be rather narrow, referring to the "expressions that advocate incitement to harm […] based upon the targets being identified with a certain social or demographic group". Many studies have proven how this definition falls well short of addressing the complexity of hate speech in the contemporary global world (Gagliardore et al. 2015). Academic debate, on the other hand, seems to have gone too far the other way, lacking agreed definitions of the concept, with implications involving "substantives such as hatred, extremism, racism, hostility and verbs like attack, wound, denigrate, silence, threaten, harass, or defame" (Holkeri 2013, 77). Also the broader policy definitions, such as those of the Council of Europe, have proved to be rather ineffective, as they are widely aimed at "covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin" (Council of Europe 1997).

---

[1] Serra Hunter Fellow, Dept of Social Anthropology, Universitat de Barcelona. Montalegre, 6 08001 Barcelona (Spain)
[2] See e.g. Awan (2013); Williams and Burnap (2015), evidencing how this was the case in the United Kingdom following the Woolwich attacks in May 2013.

744

An Ethnographic Approach

Although international institutions, including the EU, have paid increasing attention to the phenomenon of online hate speech in the last decade, describing it as a global problem[3], there is a distinct lack of research in the field, especially regarding qualitative studies on the experiences of online hate speech. Furthermore, studies on online hate speech have a tendency to focus on the content, dissemination and the prosecution, from a legal perspective (see e.g. Glaser et al 2002; Brown 2009; Waldron 2012; Foxman 2013) or on specific target groups, especially in the case of islamophobia (see e.g. Ekman 2015; Awan 2014). Yet, the relation of youth and online hate material remains under-researched, particularly from the anthropological tradition of ethnography.

Both these gaps, the ethnographic approach and the focus on the youth, are the object of this paper, grounded on a multi-sited ethnographic research, under a comparative perspective[4]. Taking France, Italy, Romania, Spain and the United Kingdom as the qualitative case studies[5], this paper discusses how young people experience, reason around and react to racist and xenophobic online hate speech. The focus has been placed on the direct victims, on individuals belonging to groups targeted by hate speech; as well as on bystanders. A key concern, as argued on the basis of the empirical findings presented here, relates to how young people understand the interconnections between their online and offline worlds, and how they react to the hateful material that they encounter on web platforms. The research hypothesis, substantiated through the ethnography, stems from the problematization of hate speech as two fundamental rights in conflict: *freedom of speech*, to protect opinions stated in public and the *prohibition of discrimination*, protecting persons and collectives from behaviours which lack respect, humiliate or depreciate them in relation to certain features (Rey Mart ńez 2014).

Conceptualising Current Online Hate Speech

Online hate speech encompasses a variety of racist and xenophobic content; from that spread on specific hate sites by far right groups, to racism expressed in posts and comments on social media. Whereas hate speech expressed online is not essentially different from hate speech offline, it does entail characteristics related to the channels through which it is expressed. In this regard, understanding the specifications of technological interactions is critical to diagnose its influence. Although the conditions of social media are not in and of themselves new, their interrelations do create new opportunities and challenges (Boyd 2014). These include permanence or persistence, referring to the durability of online expressions (Gagliardone et al. 2015); visibility, which concerns the potential audience reached; searchability which relates to the possibility for content to be found; and itinerancy or "spreadability" (Boyd 2014), alluding to the ease with which content can be shared, meaning that even when hate speech is removed from one place, it may be revived somewhere else (Gagliardone et al. 2015). In addition, the transnationality of online spaces enhances the effects of hate speech, posing complications regarding legal mechanisms for combating this phenomenon. Also the research has evidenced that social media creates a feeling of anonymity and impunity by which individuals are more likely to verbalise ideas they would not express in other situations for fear of being criticized or punished. The web produces an "illusory" depersonalization in regards to reality, for both uses and perpetrators, which is precisely why most hate speech perpetrators limit themselves to "keyboard activism" and do not engage in real actions.

In spite of this, it would be a fallacy to present any online action as separate from the offline world. The empirical findings show that social media is inevitably interlinked with the physical

---

[3] See e.g. the 2015 ECRI report highlighting online hate speech as a main trend of the previous year, emphasising that "hate speech through social media is rapidly increasing and has the potential to reach a much larger audience than extremist print media were able to reach previously", and the 2015 UNESCO report "Countering online hate speech".
[4] This research is grounded on the PRISM project "Preventing, redressing and inhibiting hate speech in new media" with financial support by the Fundamental Rights and Citizenship Programme of the European Union 2014-2016.
[5] 148 in-depth interviews with professionals and young social media users were conducted for this study. The focus on the perpetrators was based on a mapping of social media xenophobic and far right parties in five countries.

world as our offline lives are influenced by social media and vice-versa. An example of what is sometimes referred to as the 'digital dualist fallacy'[6] is concerned with far right groups that use the web to recruit people, whilst not considering the sites in themselves as outlets for racism (Daniels 2009). Overcoming the digital dualist perspective means emphasising the contextualisation of online hate speech as an expression of a larger phenomenon, that is of a pervasive and increasing racism, especially in Europe and the USA. Thus, when looking into young people's exposure, reactions and responses to online hate speech it is crucial to relate these to the bigger picture because, as Daniels notes, 'young people who are immersed in digital media do not speak with a pure voice when it comes to race and racism, but rather speak with an infected voice that both mirrors and shapes the culture and institutions in which they grow up' (Daniels 2008, 148).

As the current research shows, exposure to risks is higher among the most active users, often referred to as 'digital natives' (Prensky 2001), to whom the internet can function as a tool for harassment and as a force multiplier (Daniels 2008). Simultaneously, cyberspace has enabled online racism to flourish, with extreme right wing sites moving their content from explicit hate sites to blogs and social media platforms (Rogers 2013), more accessible to a wider audience. Yet, the ethnography shows how hate speech is not only generated by organised far right groups, but produced in common online settings by isolated individuals. This underlines the need to widening the scope of the analysis of youth's exposure to online material, to include a review of how hate speech rhetoric is used in social media (Oksanen,2014).

Online racist hate speech contributes to stigmatising, marginalising, and intimidating members of distinct and vulnerable groups. Drawing on Waldron (2012) we can define two types of messages conveyed by online hate speech: one directed to the targeted group, aiming to dehumanise and diminish them; and another directed to those with similar views, reinforcing a sense of a group threatened by "the others", providing coherence to the group. In this sense, hate speech both divides and unites simultaneously (Gagliardone et al. 2015), making a distinction between who is inside and outside of the desired society or group. Immersing a user in a context of people sharing the same opinion may increase the feeling that the world is divided into two groups: us and them. Users navigating the Internet may find little that challenges their preconceptions about people of different origins, ethnicities or religions (Delgado and Stefancic 2014). In this regard, young people may be affected by online hate speech as potential or direct victims, with damaging effects at an individual, group, and societal level. Precisely the harm to the victims and target groups of hate speech is a factor often forgotten in the debates on freedom of speech, which tends to be experienced cumulatively (Gelber and McNamara 2015). On the other hand, young bystanders may be influenced by the messages, recruited by hateful ideologies, or spurred into acts of violence by online communities (Oksanen et al 2014), with potentially having very dramatic consequences offline (Daniels 2008). This is corroborated by further evidence that shows how Internet hate material may lead to hate crimes (Wolf 2004), and Internet access can increase the incidence of racialized hate crimes (Chan et al 2015).

The Context and the Narratives

Regarding the contexts, most young people's experiences are connected to specific settings or events, such as the football World Cup or beauty pageants, indicated as triggering hate speech on social media. In this sense there is a high awareness of the role of mainstream media in perpetuating negative stereotypes, with several youngsters recollecting hateful discussions in the comments sections of digital newspapers or in the posts of social media articles. Furthermore, youngsters perceive the role of 'ordinary people' using social media and in particular Facebook and Twitter to vent their dissatisfaction, and then turning this into hatred towards specific target groups. In most cases, the hate speech content seen by the young people stems from persons unknown to them, or friends of friends. In addition, youngsters reflect on their own implication, admitting to often

---

[6] Jurgenson, N. 2011. Digital Dualism and the Fallacy of Web Objectivity:
http://thesocietypages.org/cyborgology/2011/09/13/digital-dualism-and-the-fallacy-of-web-objectivity/

reproducing or laughing at racist jokes 'in private'. The public face of racism is very different from its private expression, as privacy in a conversation seems to grant talkers permission to express ideas that they tactically avoid in public (Myers and Williamson 2001). In this sense the research evidenced that hate speech on social media also occurs as part of ordinary conversations in the shape of comments or 'likes' on news pieces or images or even framed as jokes posted by others. The great majority of youngsters interviewed have witnessed or experienced what could be broadly defined as racist hate speech[7] on social media.[8] Some of this was detailed by bystanders having seen hateful posts directed towards target groups, such as pictures targeting Afro-Americans, Muslims or Roma communities, whereas a number of young persons with minority or migrant backgrounds had experienced hateful attacks directed towards them personally.

The recollection of hateful phrases is more common among the direct targets or victims, whereas the bystanders tended to recall only the general target group or the context. However, not all young people that have witnessed hate speech were able to identify specific topics or targets. Furthermore, despite the general awareness of the existence of racism and a generalised rejection of this, hate speech on social media tends to be a normalised phenomenon amongst both bystanders and target groups. A main reason for this is that many youngsters attribute hateful comments to people reproducing discourses surrounding them, but not as the focus of their everyday interaction. As the empirical results show, for most young people social media is primarily a tool for communicating with friends and generally follow social media profiles and pages close to their interests and values.

This is linked to how the narratives of most youngsters reflect a widespread *laissez-faire* attitude, rather of indifference to the impact of online hate speech. Such normalisation among those belonging to groups targeted by hate messages often reaches the point of considering online hate speech as inevitable. This is especially the case in regards to islamophobia, with a number of cases where young Muslims described daily encounters with online islamophobia as a part of their life that they 'just have to deal with'. However, these 'normalised' experiences of suffering hate speech may often lead to refraining, partly or fully, from expressing their religious or cultural identity.

Reacting and Responding to Hate Speech

Reactions among the young social media users are reflected in their different responses to hateful content, which can be divided into four categories. The first and most common category of response to hateful content can be defined as passivity. This response results from a range of attitudes and perspectives related to the non-interventionist attitude towards online hate speech. A second category is that of minimum reaction, such as 'unfriending' or blocking users, and leaving groups or 'unliking' pages where offensive comments are shared, which was a response strategy expressed by a smaller number of young people. A third way of responding to hateful content is to report it to the social media platforms, through formal reporting mechanisms. Most youngsters show awareness of the possibility to report pages, posts or comments on Facebook.[9] However, despite the familiarity with this mechanism, by and large, youngsters tend not to report hate speech on social media, mainly due to disbelief that anything will be done. The fourth, less common response to online hate speech, can be referred to as counter-speech, such as participation in a discussion counteracting hateful arguments and introducing other viewpoints. Most of the young social media users in the study tend not to intervene in this way, mainly because of the previous backlash received when trying to engage with other users, or because they do not want to waste efforts on people who only make hateful comments as a way of seeking attention.

---

[7] In some of the interviews, where the young persons were not familiar with the concept of hate speech - especially in the age bracket 14-18, hate speech content and experiences were discussed in terms of "racist comments" or "offensive comments" or through a longer explanation by the interviewer on the range of content the discussion could refer to.
[8] Across the sample Facebook was the most popular social media platform, followed by Instagram and Twitter. Consequently, most of the hate speech experienced by the young were on Facebook. Regarding the extent of use of social media, youngsters reported visiting social media daily or several times per day, mainly through smartphones.
[9] In general there was less familiarity with this function on Twitter.

In terms of youngsters taking actions against hate speech online, evidence shows an immense underreporting, partially because vulnerable collectives lack trust in the law enforcement agencies. Evidence shows that victims of vulnerable collectives tend not to report directly to the police, but to NGOs which provides them advice and accompany them to the police if needed. In this sense, the police recounted having made more progress following awareness raising actions, mainly through collaborating with associations and NGOs. Also police investigations into online hate speech tends to be "reactive" and based on specific complaints by concrete victims. Thus, although there tend to be specific police units that deal with reports of online hate speech, these often deal with all other crimes committed online and do not have resources to "patrol" the internet.

Finally, another aspect of underreporting concerns the limited tools allowed by the legislation. In terms of prosecution, there is a general contrast between prosecuting cases of hate speech on standard web pages and on social media, where the servers hosting the web pages are obliged by law to facilitate information. International collaboration is also possible regarding web pages, although some experts consider the system rather inefficient. Social media companies like Twitter and Facebook, on the other hand, tend not to collaborate much, which the study shows is a major problem for online experts. This is also related to the limitless freedom of speech in the US or Canada where most of these companies are registered. This is summed up in the following words of a Spanish prosecutor: "The problem with justice is that it has borders, but today's criminality does not". The problem of tracking people is hardly technical but more bureaucratic. However, some experts point out that many posters of hate speech are no longer anonymous as they do not consider their posts illicit or are not ashamed by them,. A reflection of the current societal climate is recently evidenced by the ruthless *Vote Leave* Brexit campaign in the UK[10] and its immediate consequences.

In this context the perspective of social media providers on responding to online hate speech becomes essential. Non-media professionals express how an important obstacle to inhibiting hate speech is that digital newspapers and social media providers tend to perceive themselves as mere intermediaries, and take little responsibility for the content of the discussions. This is coupled with a lack of transparency in their reporting mechanisms, such as about how the moderation is done. As pointed out before, the slowness of reaction by Facebook and Twitter, contributes to a *laissez-faire* attitude by young media users and a perception that reporting hate content does not lead to any consequences. However, a change may be underway in terms of Facebook's management of hate content: Following calls from the German justice minister to do more to combat hateful speech about refugees (September 2015) Facebook has agreed to work with Germany's ministry of justice, internet service providers, and other social networks to form a task force aimed at flagging and removing hateful content more quickly. Facebook has also pledged to help finance organisations that track online hate speech, although it is not clear whether the company will change its policy on what is considered offensive material[11].

As a concluding remark empirical research continues to show how critical virtual engagement is in our global and heterogeneous societies, with vital consequences in the lives of young people. This underlies the need to ensure that young media users become aware of the impact of online hate speech, understanding that there are people behind the screens that produce the discourses and people that are strongly affected by them. In this regard, qualitative studies play a key role to unravel young people's attitudes, interactions and responses in relation to hate speech content on social media, as well as to give a voice to those that suffer from hate speech, revealing stories that too often remain untold. As shown in this paper, it is only through in-depth research that we can get a grasp on a profound and realistic comprehension of the experience and effects of hate speech and, in turn, contribute to realistic and effective policies and actions to counteract it.

---

[10] The National Police Chiefs' Council (NPCC) reported that in the week following the Brexit vote reported hate crimes increased by 500%. For furhter debate on racism and hate speech in the *Vote Leave* Brexit campaing see Jubany (2016)
[11] Wall Street Journal September 14, 2015 "Facebook Outlines Measures to Combat Racist and Xenophobic Content", http://www.wsj.com/articles/facebook-outlines-new-measures-to-combat-racist-and-xenophobic-content-1442242351

References

[1] Gagliardore, I., D. Gal, T. Alves and G. Martinez. 2015. *Countering online hate speech*. UNESCO Series on Internet Freedom. Paris: UNESCO Publishing.

[2] Holkeri, E. 2013. "Concepts and dimensions of online hate: A literature review". In *Forskerseminar, NSfKs 55,* edited by A. Storgaard. Aarhus: Aarhus University.

[3] Glaser, J., J. Dixit and D.P. Green. 2002. "Studying hate crime with the internet: what makes racists advocate racial violence?", in *Journal of Social Issues*, 58(1), 177-193.

[4] Brown, C. 2009. "WWW.HATE.COM: White supremacist discourse on the internet and the construction of whiteness ideology", in *Howard Journal of Communications*, 20:189-208.

[5] Waldron, J. 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press.

[6] Foxman, A. 2013. *Viral hate: Containing its spread on the Internet*. New York: Palgrave.

[7] Ekman, M. 2015. "Online Islamophobia and the politics of fear: manufacturing the green scare", in *Ethnic and Racial Studies*, 38:11. DOI:10.1080/01419870.2015.1021264

[8] Awan, I. 2014. "Islamophobia and Twitter: A Typology of Online Hate against Muslims on Social Media", in *Policy and Internet*, Vol 6, Issue 2.

[9] Rey Mart ńez, F. 2014. *Discurso de odio y racismo l íquido*

[10] Boyd, D. 2014. *The Social Lives of Networked Teens*. New Haven: Yale University Press.

[11] Daniels, J. 2009. *Cyber Racism: White Supremacy Online and the New Attack on Civil Rights*. Lanham, MD: Rowman & Littlefield Publishers.

[12] Daniels, J. 2008. "Race, Civil Rights, and Hate Speech in the Digital Era". In *Learning Race and Ethnicity: Youth and Digital Media*, edited by A. Everett. Cambridge, MA: The MIT Press.

[13] Prensky, M. 2001. "Digital natives, digital immigrants". In *On the horizon*, 9(5): 1–6.

[14] Rogers, R. 2013. *Right-wing formations in Europe and their counter-measures: An online mapping*. Amsterdam: Digital Methods Initiative.

[15] Oksanen, A., J. Hawdon, E. Holkeri, M. N äsi, and P. R äs änen. 2014." Exposure to Online Hate among Young Social Media Users". In *Soul of Society:* Emerald Group Publishing Limited.

[16] Delgado, R. and J. Stefancic. 2014. "Hate Speech in Cyberspace", in *Wake Forest Law Review*, Vol. 49; University of Alabama Legal Studies Research Paper No. 2517406.

[17] Gelber, K. and L. MacNamara. 2015. "Evidencing the harms of hate speech", in *Social Identities*, DOI: 10.1080/13504630.2015.1128810

[18] Chan, J., A. Ghose and R. Seamans. 2015. "The Internet and Racial Hate Crime: Offline Spillovers from Online Access", in *MIS Quarterly; NET Institute Working Paper No. 13-02.*

[19] Jubany, O. 2016. "Screening Asylum in a Culture of Disbelief: Truths, Denials and Skeptical Borders" Palgrave-Macmillan, UK, ISBN 978-3-319-40748-7