What is online hate speech?

Social media, such as Twitter, Facebook, Instagram and YouTube, are increasingly being used by members of the public to publish content online. Communications on social media take a number of forms, including text, image and video posts. The majority of communications on social media are mundane and harmless in nature (e.g. updates on daily activities, comments on TV shows or news stories, sharing of photos and videos of friends and family).

However, a minority of social media users post antagonistic and hateful comments aimed at individuals and wider communities. Hateful communications sent via social media targeted at individuals based on their personal characteristics, such as race, religion, sexual orientation, disability and transgender identity, are capable of amounting to criminal offences (see next section.

While not all antagonistic and/or hateful social media posts will amount to criminal offences, their harmful impacts on the targeted individual or group/community can still be significant.

The Law Pertaining to Online Hate Speech

The Police and Crown Prosecution Service (CPS) view all hate crime seriously, as it can have a profound and lasting impact on individual victims, undermining their sense of safety and security in the community. Hate crime covers offences that are aggravated by reason of hostility towards the victim's race, religion, disability, sexual orientation or transgender identity.

In England and Wales, hate crime is prosecuted under a range of legislation including the Crime and Disorder Act 1998, the Criminal Justice Act 2003, the Malicious Communications Act 1998, the Communications Act 2003, the Protection from Harassment Act 1997, the Offences Against the Person Act 1861 and the incitement provisions of Part III of the Public Order Act 1986.

Hateful social media posts (other than those which amount to specific offences in their own right such as making threats to kill, blackmail, stalking etc.) will be considered to be criminal if:

- Their content is grossly offensive
- Their content is threatening or abusive and is intended to or likely to stir up racial hatred
- Their content is threatening and is intended to stir up hatred on the grounds of religion or sexual orientation

It is important to note that when considering cases involving hateful communications, prosecutors operate a high threshold at the evidential stage and consider whether a prosecution is in the public interest based on the nature of the communication and the impact upon the targeted victim.

They must also be satisfied that the communication is not protected under the free speech principle (under Article 10) of the European Convention on Human Rights, that provides the freedom to cause offence.

Practitioners should refer to the CPS 'Guidelines on prosecuting cases involving communications sent via social media' for further details (http://www.cps.gov.uk/legal/a_to_c/communications_sent_via_social_media/ (http://www.cps.gov.uk/legal/a_to_c/communications_sent_via_social_media/)).

Patterns of Hate Speech Online

Academic research has identified various patterns of hate speech on the Internet. Levin (2002) studied how US right-wing groups promoted their goals online largely unchallenged by law enforcement, concluding that the online medium has been useful to hatemongers because it is i) economic; ii) far reaching; and iii) protected by the First Amendment in the United States of America. While free speech provisions are different in the UK, this finding is important because most special media companies are based in the US and follow the principles of the US constitution.

Perry and Olsson (2009) found that the Internet has created a new common space that fosters a 'collective identity' for previously fractured hate groups, strengthening their domestic presence in counties such as the US, Germany and Sweden. They warn a 'global racist subculture' could emerge if hate speech online is left unchallenged.

Leets (2001) in a study of the impacts of hate related web-pages found that respondents perceived the content of these sites as having an indirect but threatening impact on individuals and communities, generating anxiety and feelings of exclusion.

The 'Tell MAMA' (Measuring Anti-Muslim Attacks) project (http://tellmamauk.org (http://tellmamauk.org)) found that 74 per cent of all anti-Muslim sentiment reported on the site occurred online.

In a Finnish study, Oksanen et al. (2014) show how 67 per cent of 15 to 18 year olds had been exposed to hate material on Facebook and YouTube, with 21 per cent becoming victims of such material. This final study evidences how the rise of social media platforms has been accompanied by an increase in cyberhate.

There are several resources that track the pattern of hate speech globally. On Twitter the so called 'sentinel' sites @YesYoureRacist, @YesYoureGaycist and @YesYoureSexist track the production of racists, homophobic and sexist tweets and challenge individuals using hateful language online.

The website www.nohomophobes.com tracks homophobic tweets and shows that there have been over 50 million tweets (internationally) containing homophobic phrases since it started monitoring in 2012.

Hate Speech Watch (http://www.nohatespeechmovement.org/ (http://www.nohatespeechmovement.org/)) is a European online database set up to monitor, share and discuss hate speech content on the Internet.

Hate Speech Online and Trigger Events

Research has shown that the prevalence and severity of offline hate crimes are influenced in the short term by singular or clusters of events. Acts of terrorism have been shown to influence the prevalence of anti-immigrant sentiment and hate crimes and incidents.

Across Europe Legewie (2013) found a link between anti-immigrant sentiment and the Bali and Madrid terrorist bombings, in the US King and Sutton (2014) found link between terrorist acts and a rise in hate crime incidents, and in the UK, Hanes and Machin (2014) found significant increases in hate crimes reported to the police in London following 9/11 and 7/7.

Williams and Burnap (2015) show that following trigger events, such as terrorists acts, it is often social media users who are first to publish a reaction. In the aftermath of the killing of Lee Rigby in Woolwich, social media users identifying with right wing political groups were most likely to produce hateful content on Twitter.

Like offline hate, cyberhate was shown to spike and rapidly decline within the first 48 hours of the attack, indicating that cyberhate has a ‘half-life’. Williams and Burnap (2015) conclude that while social media can act as an amplifier for cyberhate given the significant number of people who use various platforms to spread hateful sentiment, its spread can be limited by users engaging in counter-speech (see below).

The Impacts of Online Hate Speech

The findings from the All Wales Hate Crime Project asked victims about the negative impacts of hate crimes and incidents. Hate crimes and incident can have considerable physical and/or psychological impacts on victims, their families and the wider community:

- Nearly a fifth of victims attempted to conceal their identity
- Nearly a third of victims had thoughts about moving from their local area
- One in seven hate crime victims reported having suicidal thoughts
- Victims of repeat victimisation were over four times more likely than any other victim to experience thoughts of suicide
- Being unemployed and feeling socially excluded increased the likelihood of suffering multiple types of negative impact
- Violent hate crime victims were significantly more likely to suffer negative impacts

While there currently exists limited information on the impacts of online hate on victims, some of the findings from All Wales Hate Crime Project can be applied. The project found that victims of low level, persistent hate related disorder (characteristics that are common to hate speech online) were more likely to experience the impacts of losing confidence, crying, concealing their identity, changing their appearance and retaliating verbally. These victims were also less likely to make a report to the police (see below). A report published by Tell Mama found that victims of online hate speech had experienced threats of violence, racist comments, and the creation of fake profiles for purpose of harassment. Online victims reported experiencing depression, emotional stress, anxiety and fear.

Advice for those who Encounter Online Hate Speech

All forms of online hate speech can have negative impacts on individuals, groups and communities. In some cases online hate speech may amount to a criminal offence, and when it does witnesses and victims should report it to the police.

However, not all online hate speech may amount to a criminal offence. In such cases Internet users can engage in reasonable and constructive ways to challenge hate and possibly stop it from spreading.

Reasonable and constructive ways to challenge online hate speech

Counter-speech is a common response to online hate speech. Counter speech online can have a positive effect by stemming the propagation of hate and, when involving groups of people, reinforces norms of acceptable behaviour.

Combating hate speech with counter-speech has some advantages over government and police responses: i) it can be rapid, ii) it can be adaptable to the situation; and iii) it can be used by any Internet user (e.g. members of the public, charities, the media, the police).

Researchers at Cardiff University developed the following typology of counter-speech:

- Attribution of Prejudice
  e.g “Shame on #EDL racists for taking advantage of this situation”

- Claims making and appeals to reason
  e.g. “This has nothing to do with Islam, not all Muslims are terrorists!”

- Request for information and evidence
  e.g. “How does this have anything to do with the colour of someone’s skin??”

- Insults
  e.g. “There are some cowardly racists out there!”

Initial evidence from ongoing experiments with social media data show that counter-speech is effective in stemming the length of hateful social media conversations when multiple unique counter-speech contributors engage with the hate speech producer.

However, not all counter speech is productive, and evidence shows that individuals that use insults against hate speech producers often inflame the situation, resulting in the production of further hate speech. When

engaging in counter-speech, or advising others on its use, the following principles should be followed to reduce the likelihood of the further production of hate speech:

- Avoid using insulting or hateful speech
- Make logical and consistent arguments
- Request evidence if false or suspect claims are made
- State that you will make a report to the police or third party if the hate speech continues and/or gets worse (e.g. becomes grossly offensive or includes threats)
- Encourage others to also engage in counter-speech

How to Report Online Hate Speech

The All Wales Hate Crime Project showed that victims of low level, persistent hate related disorder avoid making reports of this type of hate to the police because a) it happens so frequently that victims become accustomed to it, b) they don't think the police can do anything, c) they are often perceived as too trivial in isolation, and d) they are unsure how seriously these incidents will be taken by the police. It is likely that victims of online hate speech hold similar perceptions and behave in a similar way.

The Police and CPS view all hate crime seriously, as it can have a profound and lasting impact on individual victims, undermining their sense of safety and security in the community. Victims should always be encouraged to make a report to the police if they feel they have been targeted with online hate speech that is grossly offensive, threatening, harassing or inciting others to engage in hate related activities.

Reports can be made directly to the police or via the True Vision website (http://report-it.org.uk/wales (http://report-it.org.uk/wales)). The website is an online platform for the reporting hate crimes and provides information for victims and advocates. It contains official strategies and polices that guide police and partners about how to respond to incidences of hate, what happens when a report of hate crime is made,

personal safety tips, and organisations which can offer support. The site also offers up-to-date hate crime data and reports. In 2013 the True Vision mobile phone app was launched to support the website.

Case studies

The case studies below demonstrate that the police and CPS take seriously hate speech online. However, not all forms of online hate speech amount to criminal offences, as the Case Study 3 demonstrates.

Case Study 1:

In 2012, Liam Stacey made several hateful comments on social media towards professional footballer who had suffered a cardiac arrest on the pitch. Police were inundated with complaints as members of the public reported Stacey's comments. The first of his messages began with "LOL [laugh out loud]. F*** Muamba. He's dead!!!" A number of people took him to task for his views and he retaliated by posting a series of offensive and racist insults, some of a sexual nature, aimed at his attackers. Stacey branded people who criticised him on Twitter as "wogs" and told one to "go pick some cotton". Stacey was sentenced to 56 days in prison, charged with Racially Aggravated Section 4A of the Public Order Act 1986. This was one of the first cases involving hate on social media that went before the courts in England and Wales.

Case Study 2:

In 2014, Isabella Sorley and John Nimmo and were jailed for abusing feminist campaigner Caroline Criado-Perez. Isabella Sorley was jailed for 12 weeks and co-defendant John Nimmo was jailed for 8 weeks for threatening behavior. Isabella Sorley used Twitter to tell campaigner Criado-Perez to "f*** off and die you worthless piece of c**p", "go kill yourself" and "rape is the last of your worries". John Nimmo told Criado-Perez to "shut up b****" and "Ya not that gd looking to rape u be fine" followed by "I will find you [smiley face]". Both pleaded guilty to sending menacing tweets, admitting they were among the users of 86 separate Twitter accounts from which Criado-Perez had received abusive messages. Caroline Criado-Perez was so harmed by the abuse she received on Twitter that she had a panic button fitted in her home.

Case Study 3:

In 2012, Daniel Thomas was arrested after a homophobic message he sent about Olympic divers Tom Daley and Peter Waterfield went 'viral'. Following arrest Thomas was not prosecuted the as the Director of Public Prosecutions (DPP) decided the message was "not so grossly offensive that criminal charges need to be brought". He continued "This was, in essence, a one-off offensive Twitter message, intended for family and friends, which made its way into the public domain. It was not intended to reach Mr Daley or Mr Waterfield, it was not part of a campaign, it was not intended to incite others and Mr Thomas removed it reasonably swiftly and has expressed remorse. Before reaching a final decision in this case, Mr Daley and Mr Waterfield were consulted by the CPS and both indicated that they did not think this case needed a prosecution." The DPP concluded "Social media is a new and emerging phenomenon raising difficult issues of principle, which have to be confronted not only by prosecutors but also by others including the police, the courts and service providers. The fact that offensive remarks may not warrant a full criminal prosecution does not necessarily mean that no action should be taken."

Additional Information Sources:

True Vision website: http://www.report-it.org.uk (http://www.report-it.org.uk)
Hate crimes on and offline can be reported to the police via this website

True Vision phone app: here (http://appcat.com/app/7007/8ad9f49c87648512f5455b88a3dd8a48/#getApp)
Hate crimes on and offline can be reported via this mobile phone app

Digital Wildfire Project #TakeCareofYourDigitalSelf
video: https://www.youtube.com/watch?v=5nXaEctiVhs (https://www.youtube.com/watch?v=5nXaEctiVhs)

The video is aimed at 9 to 13 year olds who are starting to use social media.

Digital Wildfire Project ‘What makes a good digital citizen on social media?’ video: https://www.youtube.com/watch?v=kh1_7VVoq8g (https://www.youtube.com/watch?v=kh1_7VVoq8g)

The Digital Wildfire Project asked young people “What makes a good digital citizen on social media?” This video shows some of the responses.

Hate Crime – Are you thinking for yourself: https://www.youtube.com/watch?v=nexTF4_nr7c (https://www.youtube.com/watch?v=nexTF4_nr7c)

Crown Prosecution Service Hate Crime Schools Project Packs: http://www.cps.gov.uk/northwest/working_with_you/hate_crime_schools_project/ (http://www.cps.gov.uk/northwest/working_with_you/hate_crime_schools_project/

The Crown Prosecution Service, National Union of Teachers and many community groups have worked together to produce a range of resources on hate crime. Pupils from schools throughout the country helped to devise the dramatised scenarios included in the presentations. They provide starting points for discussion and are based on real life experiences of the young people who took part in the project. Classroom activities and guidance for teachers are available. They are designed to increase pupils’ understanding of hate crime and prejudice and enable them to explore ways of challenging it.

References and Further Reading

Burnap, P. and Williams, M. L. (2015) ‘Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making’, Policy & Internet 7(2), pp. 223-242.
Burnap, P. and Williams, M. L. (2016) ‘Us and them: identifying cyber hate on Twitter across multiple protected characteristics’, EPJ Data Science 5, article number: 11.
Hanes, E. and Machin, S. (2014) ‘Hate Crime in the Wake of Terror Attacks: Evidence from 7/7 and 9/11’, Journal of Contemporary Criminal Justice, 30:247-267.
King, R. D. and Sutton, G. M. (2014) ‘High Times for Hate Crimes: Explaining the Temporal Clustering of Hate Motivated Offending’, Criminology, 51:871-894.
Leets, L. (2001) ‘Responses to Internet Hate Sites: Is Speech Too Free in Cyberspace?’, Communication Law and Policy, 6:287-317.
Legewie, J. (2013) ‘Terrorist events and attitudes toward immigrants: A natural experiment’, American Journal of Sociology, 118:1199–245.
Levin, B. (2002) ‘Cyberhate: A Legal and Historical Analysis of Extremists’ Use of Computer Networks in America’, American Behavioral Scientist, 45:958-988.
Oksanen, A., Hawdon, J., Holkeri, E., Nasi, M. and Rasanen, P. (2014) ‘Exposure to Online Hate among Young Social Media Users’, in M. Nicole Warehime (ed.) Soul of Society: A Focus on the Lives of Children & Youth, 253-273. Emerald.
Perry, B. and Olsson, P. (2009) ‘Cyberhate: The Globalisation of Hate’, Information & Communications

Technology Law, 18:185-199.
Williams, M. L. 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), pp. 211-238.

(http://www.cf.ac.uk)

(http://socialdatalab.net)

(http://gov.wales/?lang=en)

(http://www.esrc.ac.uk)