DO NOT FEED THE TROLLS

kinder, gentler politics

webrootsdemocracy.org
## contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>forewords</td>
<td>3 - 6</td>
</tr>
<tr>
<td>executive summary</td>
<td>7 - 10</td>
</tr>
<tr>
<td>literature review</td>
<td>11 - 16</td>
</tr>
<tr>
<td>case study</td>
<td>17 - 20</td>
</tr>
<tr>
<td>impact and causes</td>
<td>21 - 25</td>
</tr>
<tr>
<td>policy roundtable</td>
<td>26 - 28</td>
</tr>
<tr>
<td>consultation responses</td>
<td>29 - 33</td>
</tr>
<tr>
<td>focus groups</td>
<td>34 - 37</td>
</tr>
<tr>
<td>social media analysis</td>
<td>38 - 44</td>
</tr>
<tr>
<td>recommendations</td>
<td>45 - 55</td>
</tr>
<tr>
<td>references</td>
<td>56 - 59</td>
</tr>
</tbody>
</table>

*Kindred, Gentler Politics: Tackling online abuse in political debate*
forewords

areeq chowdhury
khadija said
hannah bardell mp
magid ‘magic’ magid
Despite all of the research and all of the column inches written about online abuse, there has been no significant action taken, or progress made, with regards to tackling it. Whilst social media has great potential to be a hub of political conversation and a forum for debate, the reality is that it is fast becoming a sewer of hate speech and abusive content.

Our research has found users self-censoring their opinions online out of fear and opting instead to talk about politics in private forums. At the extreme, some told us that the abuse they received had driven them to suicidal despair. Negative political conversations online outweigh positive ones at a ratio of two to one.

If we, as a society, are to be serious about tackling this problem, we need to begin thinking about new ideas that focus on sanctions, education, and reform. That is what this report has set out to do.

Throughout this research, we received stories and ideas from victims, experts, and young people, particularly women and minority ethnic groups. Whilst to some, talk of online abuse is merely “political correctness gone mad” to others it is a strain on their mental health and an emotional struggle to be able to have an opinion without the fear of having their existence and identity attacked.

Online abuse is a spectrum with one end filled with hate speech and incitements of violence, and the other end filled with language and behaviour we would not tolerate anywhere else in society. The internet should not be treated as a special space that is above the law or beyond regulation. The losers of this system are often minorities, a group which should be empowered by such technology.

We hope that you find our report as thought-provoking and educational when reading it as we found it when writing it.

Areeq Chowdhury
Chief Executive, WebRoots Democracy

Khadija Said
Researcher, WebRoots Democracy
Hannah Bardell MP

Online abuse is a growing plague on society and is a problem which needs to be taken far more seriously by all of us. As an MP, I’ve received a range of abuse online including comments by men saying what they would like to do to me, comments about the way I look, and comments about being gay. This isn’t acceptable. Regardless of whether you’re a woman or whether you’re gay, and regardless of what race you are or what religion you practise, you should not be subjected to this kind of abuse. I therefore commend this report by WebRoots Democracy looking into how best we can tackle this issue in the future.

Social media has been a great innovation bringing us all closer in a way that was unimaginable before. I am a great champion of MPs using these platforms to reach out and connect with their constituents. However, they have also underlined how views which are racist, sexist, and homophobic are still widely held in society today. As highlighted in this study, many comments made online are, in fact, hate crimes. These crimes must be addressed by individuals, the police, and social media platforms themselves. This isn’t a new concern and collectively we have failed to tackle this problem in recent years. The proposals set out in this study deserve consideration by the Government. We must invest in education and policing, whilst applying reasonable sanctions on individuals who carry out abuse and the platforms which fail to address it.

Continued failure to combat online abuse will mean that social media platforms will never reach their true potential for civil debate and will drive away good people from getting involved with politics. As many respondents to the consultation have indicated, social media is becoming a hostile environment in which to express your political opinions online. I look forward to seeing political parties from across the spectrum coming together to tackle this.

Hannah Bardell MP
Member of Parliament for Livingston
SNP Shadow Spokesperson
(Digital, Culture, Media and Sport)
Magid ‘Magic’ Magid

Since being elected as the Lord Mayor of Sheffield I have been subjected to abuse both online and offline. This abuse systematically targets my identity as an immigrant and a proud black Muslim, three groups often marginalised and underrepresented in political debate. For far too long, online abuse has created a toxic space deterring these groups from getting involved with politics. This abuse isn’t just a tech problem, but a human one. Many of the deplorable views and comments posted on the internet are incited and inflamed by politicians who should know better. The problem is a societal one, something this report by WebRoots Democracy particularly focuses on.

We should always be free to call out bad politics and bad politicians using colourful language, as I have done myself on a number of occasions, but when this language crosses into hate speech action needs to be taken. The internet should be a free space of conversation, especially for minorities whose voices are often marginalised by the mainstream media. The ideas set out in this report could go a long way to ensuring that the internet truly is a space for civil conversation and to rooting out the underlying causes of bigoted opinions.

I was pleased to help promote the open consultation that fed into this report and, whilst alarmed at some of the stories of abuse individuals have been subjected to, I am very grateful to everyone who took the time to share their thoughts, ideas, and experiences. If we are to bring about a kinder, gentler politics, we need to see a collective effort by social media companies, politicians, and communities to root out bigotry, reform abusers, and protect victims.

Magid ‘Magic’ Magid
Councillor for Broomhill & Sharrow Vale
Lord Mayor of Sheffield, Green Party
executive summary

purpose and background
acknowledgements
recommendations
key findings
executive summary

Purpose and background

According to Hansard, the question of online abuse was first raised six years ago, in 2012.\textsuperscript{1} Since then, very little has changed, the problem has become worse, and institutions have been slow to react. The rise of online abuse has poisoned political debate and threatens to turn good people away from getting involved with politics. Whilst many politicians and public figures have spoken out against the problem, responsibility is often deflected to social media companies based overseas.\textsuperscript{2}

We embarked on this project in order to better understand the problem, particularly with regards to its impact on young people, and to set out a range of bold proposals for how governments, society, and social media companies can bring about significant changes to enable social media platforms to become civil spaces of discussion.

The report looks at the key debates and causes of online abuse in political debate, as well as the extent and impact of it within the UK. It questions whether online abuse affects individuals’ ability to freely express political opinions online without fear of being subjected to hate speech and whether this is, in turn, making involvement with politics online less appealing.

As part of this research, we conducted an open consultation in which we received more than a hundred responses from victims and non-victims sharing their thoughts, suggestions and personal experiences. In addition, we organised an expert roundtable as well as a number of focus groups with young people, in particular women and ethnic minorities. We also investigated various political Facebook groups and analysed more than 53,000 tweets directed at leading UK political influencers, both manually and using sentiment analysis software. Finally, we built on the learnings gained from a number of public seminars we have hosted with computer science academics, technologists, and anti-discrimination campaigners.

This project has been supported by the Andrew Wainwright Reform Trust and Royal Holloway, University of London. It is the first report in a series we hope to undertake exploring potential state regulation of social media platforms.

WebRoots Democracy

WebRoots Democracy is a youth-led think tank focused on the intersection of technology and democratic participation. Based in Tech City, London, its aim is to modernise, enhance, and future-proof democracy in the UK.

With the advice of industry experts and academics, and the support of a cross-party group of MPs and Peers, WebRoots Democracy intends to deliver evidence-based tech policy to sustain the future of British democratic participation.

Acknowledgements

This report has been co-authored by a number of people, including lead authors Areeq Chowdhury (Chief Executive) and Khadija Said (Researcher). Other key contributors are Rachel Fielden; Maya Fryer; Rich Mason; Fahmida Rahman; Dr James Sloam; and Tess Woolfenden. We are also grateful to Rachel Blair; Laura Deslandes; Ben Pearson; and Georgia Wild for assisting with the research activities involved with this report.

Special thanks go to everyone who took the time to feed into our report in person, over the phone, and online.

Recommendations

Our recommendations are centred around four key themes of sanctions; reform; oversight; and education.
These are explained in depth later on in the report in the “Recommendations” chapter but are listed and summarised below.

**Civil Internet Tax**
*Categories: Sanctions; Reform; Oversight; and Education.*

The Civil Internet Tax would be a new tax levied on social media platforms to raise money to fund digital literacy and anti-discrimination initiatives as well as provide resources to help police forces address cases of online abuse. In addition, it would support the creation of a new regulatory body for social media platforms called Ofsoc (Office for Social Media Regulation).

**Ofsoc (Office for Social Media Regulation)**
*Categories: Sanctions and Oversight.*

The Office for Social Media Regulation would be a regulatory body monitoring the practices of social media companies. The remit of Ofsoc may include overseeing how effective social media companies are in combating online abuse and to collate and analyse quarterly transparency reports. Ofsoc would also be a body that users could write to and raise complaints about abuse received on social media platforms, akin to the role of Ofcom and broadcasting complaints.

**Pseudonymous accounts**
*Categories: Reform and Oversight.*

This proposal is for social media platforms to consider how best they can incentivise users to verify their identities and privilege those that do. The aim would be for social media users to have the option of being anonymous on the front end of the platform but for social media companies to know who their users are and whether they are real people as opposed to bots. This recommendation also calls on the Government to commission research into the links between anonymity and online abuse.

**Quarterly transparency reports**
*Categories: Oversight.*

This would be a requirement for social media companies to submit quarterly transparency reports to Ofsoc including, amongst other things, detail on the progress made in tackling online abuse and hate speech on their platforms. They would also provide data on the number and types of reports of abuse received over the preceding quarter.

**Nationwide advertising campaign**
*Categories: Reform and Education.*

To help tackle the underlying issues behind online abuse, this proposal is for a nationwide anti-discrimination and digital literacy advertising campaign. The campaign would take place both online and offline. Akin to a public health campaign, the primary audience would be adult social media users in the UK.

**Investment in anti-discrimination campaigns**
*Categories: Reform and Education.*

This proposal would earmark 25% of the annual revenues raised from the new Civil Internet Tax for investment into anti-discrimination initiatives in the UK. Similar to the nationwide advertising campaign, this recommendation represents a long-term investment in bringing communities together and combat some of the underlying root causes of abuse, both online and offline, that occur throughout society.

**Platform suspension powers**
*Categories: Sanctions.*

Platform suspension powers would act as a final sanction to be taken by the state to reprimand social media companies that consistently fail to significantly tackle hate speech on their platforms. This would be a radical reform requiring the input of Parliament, the judiciary, Ofsoc, and internet service providers. The powers would block sanctioned social media websites from operating in the UK for up to three consecutive days.

**Digital ASBOs**
*Categories: Sanctions, Reform, and Education.*

Digital Anti-Social Behaviour Orders (DASBOs) would be a reincarnation of the ASBOs that were
in place in the UK from 1998 to 2015, and still used in Scotland. The DASBO would operate in a similar manner to the Community Protection Notice (CPN) that replaced ASBOs in England, Wales, and Northern Ireland. The DASBO would prevent offenders from using certain social media platforms for a specific time period. In addition, those with DASBOs would potentially be placed onto a publicly accessible “online abusers register” (described later in this chapter). DASBOs would be preceded by a written warning and would be issued for less serious cases of online abuse.

**Online abusers register**

*Categories: Sanctions, Reform, and Education.*

The Online Abusers Register would be a publicly accessible register of individuals that have received a DASBO for persistently engaging in abusive behaviour online. Employers, in particular, would be encouraged to use the register during their recruitment process. Details of the user, the nature of the abuse, and the precise quote of abuse would be listed on the register. To encourage reform, an individual’s name would only be published on the register after a three-month period and can be prevented if the offender attends classes on anti-discrimination and digital literacy. The aim of the register, therefore, would be to act as a disincentive to commit online abuse, protect citizens, and to encourage reform of the offender.

**Online Neighbourhood Watch**

*Categories: Oversight.*

The Online Neighbourhood Watch would be a scheme promoted by social media companies to incentivise users to report the abuse they see on the platform. The reporting of abuse would be encouraged through a rewards system integrated into the platform itself.

**Innovative T&Cs**

*Categories: Education.*

This proposal calls on social media platforms to innovate the method they display and obtain consent from users to their terms, conditions, and community standards. A number of major social media platforms already have strong sets of community standards; however, these are not made readily or prominently available to new and existing users. This recommendation calls on platforms to utilise their skills in creativity and innovation to ensure that all users, including teenagers as young as 13-years-old, can provide informed consent to behaving in a civil manner online.

**Key findings**

- Online abuse can affect all groups of people who engage in political conversations online, however women and minorities, in particular, are often subjected to a toxic level of abuse which focuses on their identities.
- Enforcement of existing hate speech laws is seen as insufficient despite the definition equally applying to verbal abuse online as well as verbal abuse offline.
- Respondents to our open, online consultation identified social media companies as having the most responsibility for tackling online abuse, followed by abusers themselves, and the police.
- The rise of online abuse may be affecting how we talk about politics online with a number of the young people that engaged with this research stating that they are increasingly expressing political opinions in private groups and forums rather than in public ones. In addition, a number of young people told us they self-censor political content they post online in order to avoid being subjected to abuse.
- According to our manual sentiment analysis of thousands of tweets directed at UK-based political influencers, the majority of conversations are neutral in nature (66%). However, negative interactions outweigh positive interactions at a ratio of 2:1.
- A number of popular UK political Facebook groups contain abusive content and content which could be classed as “hate speech”. One group in particular (with 7,000 members) contained numerous examples of racism and Islamophobia, as well as an incitement to assassinate the Mayor of London posted by the group moderator.
literature review

defining hate crimes online
responsibility

global examples
literature review

Having an online presence has become integral to the role of MPs, political influencers and journalists in the UK today. It allows for a quick dissemination of messages, a space promoting political engagement and a platform providing another avenue to hold elected officials and influencers accountable to the claims they make. However, with the rise of online abuse on social media platforms it has become increasingly clear that users are not always held accountable for the abuse they direct at politicians and political influencers. The efficiency and speed in which users can tweet has also meant that the volume of abuse is harder to track and manage.

The situation has become so toxic that in July 2018, Parliament’s health and wellbeing service advised MPs to close down their Twitter accounts, particularly female MPs that are routinely targeted with threats of violence. They found that a high number of MPs seeking help for their mental health cited social media as a factor contributing to their illness. Research has shown that an “overwhelming majority” of female MPs have received online and verbal abuse from the public, and a third have considered quitting as a result.3

There has been a range of research undertaken on the issue of online abuse in the UK and abroad. In particular, studies have focused on abuse targeted towards elected politicians such as MPs. Whilst our paper is focused more widely on abuse within political debate, particularly towards young people and influencers, we have drawn upon major studies from the likes of Amnesty International4 as well as research from the United Nations5 and the House of Commons Library6 to help inform our recommendations. In this report, we will build on previous studies and explore the impact of online abuse on democratic engagement and whether or not it negatively impacts an individual’s desire to express political opinions online or to engage more formally with politics in the UK.

Our review covers some of the existing literature on definitions of hate speech and free speech, the arguments over where responsibility for action lies, and highlights global examples of varying methods of regulation.

The three studies we have particularly explored are Amnesty International’s report, Toxic Twitter, Atalanta’s (Anti)Social Media7, and the Home Affairs Select Committee’s report Hate Crime: abuse, hate and extremism online8. We have also examined a number of government publications and work undertaken by the Law Commission9 on online communications. What is clear from existing research is that the issue of online abuse relies heavily on the parameters drawn between free speech and hate speech. To be able to discuss this effectively in the course of this report, these terms must first be defined.

Defining hate crimes online

Free Speech

Free speech is defined in Article 10 of the Human Rights Act (1998) as the ‘freedom to hold opinions and to receive and impart information and ideas without interference by public authority’ and that under UK law ‘everyone’ has the right to this.10 Amnesty International builds on this definition by outlining the importance of responsibility with free speech and governments’ obligations to ‘prohibit hate speech and incitement’ when free speech is taken advantage of. Article 19 of the International Covenant on Civil and Political Rights11 (ICCPR) restricts free speech on anything that lacks ‘respect of the rights or reputations of others’ and ‘for protection of national security.’

Hate Speech

There is no specific mention of ‘hate speech’ under UK law. However, the Public Order Act (1986) has been cited as a means to define hate speech as the use of ‘abusive or insulting words or behaviour’ with the intent to harass, alarm or distress.12 Furthermore, it is an offence under the UK Racial and Religious Hatred Act (2006) to use ‘threatening words or behaviour, or display any written material which is threatening’ against a person targeting their religion or race.13 The same laws for hate speech and hate crime for cases
offline are also applied online under UK law. As summarised in a report by the Foreign and Commonwealth Office, freedom of expression is to be protected, but ‘advocacy of hatred is to be prohibited.’

The Home Office defines ‘hate crime’ as ‘any criminal offence which is perceived, by the victim or any other person, to be motivated by hostility or prejudice based on a personal characteristic’ and that it can be motivated by ‘disability, gender identity, race, religion or faith and sexual orientation.’ During an announcement of a new online hate crime hub, the Head of the Crown Prosecution Service said of hate speech online that “whether shouted in their face on the street, daubed on their wall or tweeted into their living room, the impact of hateful abuse can be equally devastating.” The then Home Secretary, Amber Rudd, also sought to draw an equivalence between offline and online by stating that “what is illegal offline is illegal online, and those who commit these cowardly crimes should be met with the full force of the law.”

Responsibility

The UN Guiding Principles on Business and Human Rights requires that businesses ‘know and show that they respect human rights in practice’ as part of their responsibility to protect human rights. Showing is defined as ‘providing a measure of transparency and accountability to individuals or groups who may be impacted and to other relevant stakeholders, including investors.’ In practice, this would mean that procedures to protect users from hate speech as well as the terms and conditions preventing hate speech must be made clear to all users of a platform.

Amnesty International places a large measure of responsibility on platforms and corporations to tackle online abuse. In their Toxic Twitter report they argue that there should be an ‘ongoing and proactive’ assessment of ‘how its policies and practices impact on users’ rights to non-discrimination, freedom of expression and opinion, as well as other rights, and take steps to mitigate or prevent any possible negative impacts.’

Facebook, Microsoft, Twitter and YouTube all follow the Code of Conduct on Countering Illegal Hate Speech Online under the European Commission. This requires companies to lead ‘in countering the spread of illegal hate speech online’, show ‘clear and effective processes’ to review notifications regarding illegal hate speech and share best practice with other internet companies, platforms and social media companies. However, the vague terms used in the code of conduct doesn’t bind these companies to any specific targets. The Home Affairs Select Committee in their 2017 Hate Crime report criticised social media companies’ reliance on users to report content as ‘outsourcing the vast bulk of their safeguarding responsibilities at zero expense’. The report recommended that ‘all social media companies introduce clear and well-funded arrangements for proactively identifying and removing illegal content.’ The report also suggested that the police should be able to recover the costs of enforcement from those companies.

Furthermore, the Home Affairs Select Committee and Amnesty International have called for more transparency from social media companies. The report says it is ‘unacceptable’ that Facebook and Twitter refuse to disclose the number of people they employ to safeguard users or the amount that they spend on public safety initiatives because of ‘commercial sensitivity’.

“These companies are making substantial profits at the same time as hosting illegal and often dangerous material; and then relying on taxpayers to pay for the consequences. These companies wield enormous power and influence and that means that such matters are in the public interest.”

Hate Crime Report

Atalanta’s (Anti)Social Media study highlights that with regards to sexist abuse online, the responsibility to tackle and prevent abuse fell on women themselves. Whilst they note the merit in actions which can be taken by victims themselves, they also point to male public figures adding ‘fuel to the fire’ by promoting sexism themselves. In particular, they point towards comments made by US President, Donald Trump. As for the social media platforms, they argued that responsibility
for illegal content should be shifted to the likes of Facebook and Twitter. This feeds into a deeply divided debate throughout much of the literature on this topic about whether or not social media websites are platforms or publishers.

Many studies have underlined the need for the police to act more in cases of online abuse and to be given additional resources to help them effectively tackle the issue. The Crown Prosecution Service guidelines on online crime set out a number of social media offences which include:

- **Trolling** - a form of baiting online which involves sending abusive and hurtful comments across all social media platforms.
- **Online threats** - can take many forms including threats to kill, harm or to commit an offence against a person, group of people organisation.
- **Online harassment** - can include repeated attempts to impose unwanted communications or contact in a manner that could be expected to cause distress or fear.
- **Virtual mobbing** - takes place when a number of individuals use social media or messaging to make comments to or about another individual, usually because they are opposed to that person’s opinions.

A 2018 Channel 4 Factcheck argues that existing legislation covers ‘almost all aspects of online abuse’ including laws against sending ‘grossly offensive’ messages online. In particular, the Communications Act 2003 and the Malicious Communications Act are pointed to as examples of legislation which rule against sending ‘a message or other matter that is grossly offensive or of an indecent, obscene or menacing character.’ It also makes it illegal to ‘persistently’ use social media ‘for the purpose of causing annoyance, inconvenience or needless anxiety to another.’

This adds to the voices who support more enforcement of existing laws and resources for the justice system to prosecute individuals. Some, however, have been critical of existing laws arguing that they are too punitive for the crimes committed. After student, Liam Stacey, was handed a two-month prison sentence for racist tweets sent about footballer Fabrice Muamba, the then European Commissioner for Human Rights, Thomas Hammarberg, described the punishment as excessive saying that to put him in prison was “wrong”. As well as the responsibility, the response itself is still the subject of much debate.

The recommendations of this report will focus heavily on the question of where responsibility lies, and how it should be approached. We hope to add to existing literature on the subject, the case for bolder action constrained by democratic controls.

### Global examples

To provide some context on social media regulation applied in practice, this report has highlighted four key examples of ways in which different countries regulate speech online. These examples include individual case studies illustrating the need for laws that are specific to hate speech and abuse online.

#### China

Although China has the largest internet community in the world, there is still no specific law to counteract cyberbullying. For more than a decade there have been “human flesh search engines” by which internet users team together to find and publish the personal details of Chinese residents involved in public scandals. A case of this concerns 14 year-old Ding Jinhao who wrote “Ding Jinhao was here” on the walls of an ancient temple in Luxor, Egypt. The human flesh search groups found and published his personal details, school address and even hacked his school’s website.

China has passed new cybersecurity laws requiring instant messaging services and internet companies to make it a requirement for users to register with their real names and personal information and to censor content that is “prohibited”. The terms laid out in the legislation are vague and the restriction on online anonymity has been raised as a concern by Human Rights Watch. Sophie Richardson, the China director for Human Rights Watch said “the law will effectively put China’s Internet companies, and hundreds of millions of Internet users, under greater state control.” Whilst keeping records of online identities could crack
down on online abuse and harassment efficiently, the lack of laws prohibiting online abuse and harassment offers no consolation to victims of online abuse.

**United States**

The United States has no clear law prohibiting online hate speech and hate crime. Many journalists reporting online abuse to police have had to resort to filing a civil protection order to protect themselves from stalking and harassment that stems from online abuse.

One such journalist is Amanda Hess who was contacted by her harasser soon after the civil protection order had expired. This particular case has highlighted how the lack of legislation targeting online hate and abuse, specifically, has meant that abuse is only taken seriously when the abuse is enacted physically.

**Colombia**

Gendered online abuse is particularly prevalent in this example. Many cases of online violence in Colombia target women and women’s rights campaigners specifically for their activism. At least three women working for feminist organisation Mujeres were victims of sexual violence, harassment and stalking which they believed was directly related to the online abuse they received. In 2008, Colombia introduced legislation addressing violence against women but with no mention of violence enabled by technologies making it “ill-equipped” to help victims of online harassment and abuse.

As mentioned earlier in this chapter, online abuse often targets female activists and politicians to put them off engaging in public debate. Without specific regulation and law protecting women online, online abusers are protected making social media platforms a potentially dangerous and toxic space for female activists.

**Germany**

In Europe, Germany is regularly cited as the main example of a government applying severe measures to counteract online abuse. Germany passed a law that could fine social media companies up to 50 million euros if they fail to remove hate speech, fake news and illegal material within 24 hours of it being reported. The controversial law has received widespread criticism over concerns of it stifling free speech. Human Rights Watch said the law “can lead to unaccountable, overbroad censorship and should be promptly reversed” and that the law sets a dangerous precedent for other governments to outsource responsibility in tackling hate speech. Furthermore, the swift removal of illegal content could leave users with “no judicial oversight or right to appeal” which is problematic for cases that should be taken to court.

From the German government’s point of view, the issue comes down to being able to provide a strong incentive for social media companies to take threats and online abuse seriously, and promptly. Criticism aside, this case study is an example of the government of a European nation, similar to the UK, taking responsibility for how social media companies operate in their country and treating threats made online as seriously as they would be if they were made offline.

**Conclusion**

The main arguments surrounding this topic are relatively consistent across existing literature, primarily the need to be clear about what is and is not acceptable speech both offline and online, the question of where responsibility for policing conversations should lie, and the proportionality of any sanction on those who violate these rules.

Whilst a lot of UK-specific papers on this subject have centred on the abuse directed at elected politicians, our paper sets out to explore the types of abusive content directed in normal day-to-day political conversations, in particular towards political influencers and young people.

As a number of studies have attempted sentiment analysis on social media platforms, this is something we have explored, too, however, as will be detailed later in the paper, we identify a number of issues with this method of analysis and is something we believe should be explored in future.

Our research will highlight some of the low levels that online political conversations can sink to on Twitter and in Facebook groups, as well as shine a light on the impact rising online abuse is having
on youth political engagement online. Finally, the report will make the case for bold, democratic intervention with social media giants with regards to taxation, anti-discrimination, and sanctions for failure to combat the problem.
case study

lily madigan

“If current MPs don’t act on this, young people will when we’re in power.”
Lily Madigan

Lily Madigan is a prominent LGBTQ+ rights activist and is the first transgender woman to be elected as Women’s Officer within the Labour Party. Her election, as a 19-year-old, in November 2017 created a stir online and in the media with a number of news articles criticising the allowance of transgender women to work in women-exclusive roles. In addition, her application to the Jo Cox Women in Leadership programme equally drew criticism online. This has made Madigan a target for transphobic online abuse, something she has spoken out against in the past.

This has led to a number of high-profile public figures such as Ed Miliband and Owen Jones expressing solidarity with Lily, condemning the criticisms as transphobic. She has also recently been featured in Teen Vogue’s 21 under 21 list.

In September 2018, we invited Lily to TechHub Moorgate, to be interviewed as a case study for our report and to share her thoughts and experiences about the large amount of online abuse she receives on a regular basis. Ironically, after we tweeted about the interview afterwards, our Twitter account was barraged by a number of transphobic comments, many of which looked targeted. We will detail examples of this later in the chapter.

Her experience

Lily began receiving abuse online when she came out as transgender. Her involvement with local politics and the subsequent national conversation afterwards, severely increased her vulnerability to online hate. Across her social media platforms, Lily noted that Twitter was the worst social media platform for abuse due to the inability to delete individual comments. She stated her preference for other platforms as, even though she may still be subjected to abuse on them, there was no risk of those comments being indefinitely published online.

We asked her whether it was the case that she receives a larger quantity, or a greater toxicity of abuse on privately online, via direct messaging or email, as can be the case with MPs, however she stated that the majority of abuse she receives has been public on spaces such as comment sections and tweeted responses.

“I think they are seeking validation by doing the abusing in public rather than directly messaging me in private.”

The effect

As time goes on, the effect that online abuse has upon Lily has been reduced and she now is able to ignore the majority of comments. It was clear, that abusive, transphobic comments are an occurrence which she has become normalised to.

“It says a lot of sad stuff that it’s normalised to me.”

This normalisation is something a number of politicians speak of including Bradford MP, Naz Shah, who spoke to us in detail about it in a video interview with us for this project earlier this year.

Lily mentioned however, that occasionally some instances of abuse catches her off guard and saddens her. She specifically referred to a recent incident of abuse which targeted the fact she is queer rather than attacking her transgender identity. The abuser stated:

“If she was a transgender woman and fancied women then she’s no more than a heterosexual male.”

Platform responses

We later asked Lily about the differing responses from social media platforms and whether she thought about reaching out to organisations who help victims of online abuse. She highlighted Facebook as being the worst at dealing with online abuse and that when they do respond, it is done slowly. Twitter, she argues, is slightly better and Instagram is the most effective by quickly
responding to her reports and acting in a more “liberal way”. However, she said that one of the main problems with Twitter is that it allows users to create multiple accounts hence once an abuser has been blocked it is easy for them to simply make a new account and continue tweeting abuse. She believes there should be a process put in place in order to prevent people from creating numerous anonymous accounts online to counter this.

We also enquired about her experience with the Labour Party and how the organisation has dealt with the abuse she has been subjected to:

“Labour doesn’t have a good compliance unit for dealing with so many people – that’s the main problem, however if I do report someone I do usually hear back and a few of the worst people have been suspended.”

Responsibility

With regards to where the responsibility for tackling the problem lies, Lily argues that it is “definitely” with the social media platforms. On the debate surrounding the line between free speech and hate speech, which some argue is blurred on online platforms, Lily said:

“It’s free speech up until the point where you’re affecting my free speech and I’m scared.”

Demographics of abusers

We asked Lily whether she had noticed any trends in the demographic of the abusers who have targeted her in the past. On this, she told us:

“It’s a generational thing, there are bad cookies everywhere, but by and large old men are the abusers.”

We also asked about her experience with abuse from anonymous accounts to which she explained she experiences “tonnes of it.” This is an issue we focus on in our recommendations, explained in detail later on in the report.

Sanctions

In terms of how abusers should be dealt with, Lily is of the view that culprits should be “blocked permanently from online platforms.” She also believes that working alongside the police should also be a priority for social media platforms. Lily explained that she has reported a number of online threats she has received to the police.

“In a corporate sort of sense, financial penalties are the best way to get people to cooperate.”

The future

Lily has faith that the situation will improve over time as the older generation will be replaced with a younger, more social media savvy demographic.

“If the current MPs don’t do make change with regard to this, young people will when we’re in power.”

She holds a strong opinion that representation in politics is still a big issue but one that is improving. She explained that it is always harder to go for something “if you don’t see someone like you doing it as well” adding that “hopefully other people will see me and say that if she can do it so can I.”

Online backlash

Following our interview with Lily, we tweeted from our account (@WebRootsUK) the following:

“It was so great to interview transgender activist, Lily Madigan, today. Her story will feature as a case study in our upcoming “Kinder, Gentler Politics” report, exploring the rise of online abuse in political debate.”

The tweet was subsequently targeted by numerous accounts criticising her inclusion with some descending into sexist and transphobic abuse. Lily said she had already blocked many of these accounts due to previous tweets they had sent and was unable to see the responses. Whilst ironic in its nature, the tweets underline the type of abusive content that is directed to Lily, and others like her, on a regular basis.

The tweet received 65 responses, representing an abnormally high number of responses for our Twitter account. Some examples of the tweets are listed below.
“What I find quite perplexing here is this individual is 100% physically male, he hasn’t even started hormone treatment so technically he is not transgender and certainly not transsexual. But more concerning is the repeated FAKE NEWS being whipped up by the press Re: these MEN.” - @britannia_great

“How would HE know? HE blocked everyone, especially women. Nasty little boy & here you are given HIM more space to play the victim.” - @moranrisin

“He is a misogynist women’s officer who only promotes trans issues. He is a straight man, he’s not female, he is not a transsexual.” - @vodka660

“Liam is a bloke, a woman hating blake.” - @adormidera2

“Has he still not learned to use shampoo?” - @ValerianaOlmos

“Ummm that’s a man, baby!” - @Martenlimi

“Why are you referring to someone as ‘she’ just because they put a dress on? He has a swinging cock under that pinafore. How offensive do you think that is to women?” - @SisterClaireXX

“This is a young man in his early twenties who is straight with a penis and long hair. He occasionally dresses in men’s clothing when he feels uncomfortable in women’s clothes. This man is being used as a political puppet to target progressives for votes.” - @XshudderX

“I’d rather Nigel Farage representing women in the Labour Party, at least with Nigel what you see is what you get – he knows he can be an arse and would likely admit to it. This one is a god damn snake.” - @Titus_Titus_

“I am blocked by Mr Madigan. Can someone tell him to wash his bloody hair” - @ohdofuckoffdear

“And let him know that when your knees are thicker than your thighs, it’s time to start eating cake again.” - @robjessel

As can be seen, a number of these comments actively misgender Lily, as well as dead-naming her. Other comments focus on her appearance. The overwhelming majority of these tweets were sent by accounts which do not have real names or real profile photos. In addition, in one tweet, our researcher, Maya, who was also pictured in the original tweet was compared to a “handmaid” which are women who are subjected to ritualised rape by their male masters in the popular television series “The Handmaid’s Tale.”

“That awkward moment when a report about online abuse is replied to almost exclusively with online abuse.”

A small number of users pointed out the irony of a post about online abuse, being subjected to online abuse.
impact and causes

trends in online abuse
the impact
explanations for rising online abuse
Trends in online abuse

Online abuse of individuals engaged in the UK political sphere has become a common occurrence, especially for politicians, political journalists and activists. The type of abuse received varies in form and severity from trolling, harassment and discrimination, to hate speech, doxxing and threats of violence, and will often take place on sites such as Facebook and Twitter. While awareness of the issue is growing, it appears to have done little to prevent online abuse which has at the same time been increasing in volume and severity for victims.

It is generally accepted that most individuals in the political sphere experience online abuse in some form, however not every individual or group is targeted in the same way. The evidence available demonstrates that, much like in the offline world, individuals and groups are being targeted based on their intersecting identities, including gender, ethnicity, religion, sexual orientation and health status.

Gender

According to the Inter-Parliamentary Union, social media platforms such as Facebook and Twitter have become one of the primary locations in which abuse against female parliamentarians in perpetuated. A recent study conducted by Atalanta revealed that while male and female politicians receive similar levels of derogatory comments overall online, female politicians were three times more likely to receive abusive comments related to their gender. A significant number of these comments were linked to female politician’s physical appearance, family life and relationship status, whereas male counterparts were attacked predominantly for their policies and political approaches.

A similar study conducted by The University of Salford found that female politicians receive significantly more hate speech than their male counterparts on Twitter. Of the tweets that were recorded as containing hate speech, 86% were directed at women. The report linked this to the high levels of misogyny online as well as the greater presence of men in the political sphere.

Ethnicity

Ethnicity also has a major impact on the volume and severity of online abuse received by victims. A study conducted by Evolution AI which analysed three million tweets across a twelve-month period, found that politicians from an ethnic minority receive, on average, 15% more toxic tweets than white counterparts. Amnesty International UK also found similar results when analysing the tweets received by 177 MPs in the run up to the 2017 election. They found that the 20 Black, Asian and Minority Ethnic (BAME) MPs included in the analysis received 41% of all abusive tweets, despite there being nearly eight times more white MPs included in the study. Of the abusive tweets recorded, nearly one third of them were directed at Diane Abbott, the UK’s first black female MP. The abuse she received during this period tended to focus on her gender, race and physical appearance:

“This fat retarded black bitch thinks you should be forced to feed and house a bunch of violence foreign invaders. I strongly disagree.”

“Piss off you disgusting useless fat bitch! You’re a parasite alien looking to silence native people for your power.”

Other identities

Abuse also often targets politicians’ religion and sexuality. For example, Luciana Berger MP recently became victim to a string of anti-Semitic abuse following the arrest of a white supremacist who had been abusing her online. Naz Shah MP has also faced growing Islamophobic abuse following controversial comments she made on Twitter. As reported in the Business Insider, many of the comments criticising her crossed the line into racism.

“So you support necklacing? Another corrupt muslim MP I suppose. Does the mosque tell you what you can and can’t do too? Of course it does. I think you should resign!”

Scottish Parliamentarian Ruth Davidson has also spoken out about the homophobic abuse she receives via social media.
These trends are not only visible for politicians, but also for political and human rights activists and political journalists who are facing an increase in misogynistic, violent and abusive comments. The comments below outline some of the abuse experienced by journalists and activists.

“I’ve never had abuse only because I’m a woman – it’s almost always had to do with my race.”
Charlie Brinkhurst-Cuff, journalist

“I have faced multiple layers of abuse on social media platforms, both on Twitter and on Facebook, both because I’m a woman and because of the stand I take on certain issues.”
Dawn Purvis, reproductive rights activist

“I am from a Scottish Asian community. I am a Muslim. And I’m a woman. So it’s everything. It has an exponential effect, so people will pile on the abuse for a variety of different reasons. Some of them because you are all of these things, and some because you are one of these things, or two of these things, which makes it so much more difficult to deal with, because you just wonder where do I start with this?”
Tasmina Ahmed-Sheikh, former MP

It is vital to note that the analysis presented above is not exhaustive, there are many other intersecting identities that can and should be included in the analysis of social media abuse. To an extent, we are limited by the data already collected and made available through other research. However, what is clear is that the volume and severity of online abuse received by those in the political sphere in the UK is directly affected by identities. Furthermore, the more identities you have, the more likely you are to be targeted.

The impact

The online abuse of those actively engaged in the political sphere can have serious negative consequences for both the health, wellbeing and security of victims and their families, and for democracy and political representation more generally. The below details what some of these impacts are.

Health and autonomy

Online abuse can have a serious impact on victims’ health and autonomy. As explained in a report by the European Women’s Lobby, the impacts of online abuse are no different from bullying, harassment and stalking in the offline world. This can result in “stress disorders and trauma, anxiety, sleep-disturbances, depression, and physical pain.” In an Ipsos MORI poll which looked at the experiences of online abuse for women between the ages of 18 and 55 in a number of different countries, 55% of those interviewed said they had experienced anxiety, stress or panic attacks, and 61% said they had experienced a loss of confidence as a result of abuse. These impacts can also affect the families, friends and colleagues of victims.

Security

Another serious concern for victims of social media abuse is the fear of physical safety for themselves and their families. Many comments made towards those in the political sphere are aggressive and threatening in nature, especially towards women. In the same Ipsos MORI poll mentioned above, 41% of women who had experiences of online abuse reported receiving comments that threatened their physical safety on at least one occasion. The experiences of politicians such as Diane Abbott further support this:

“I’ve had death threats, I’ve had people tweeting that I should be hung if ‘they could find a tree big enough to take the fat bitch’s weight’... I’ve had rape threats.”

The same threats are also experienced by journalists and activists. As Dunja Mijatović, former Organisation for Security and Co-operation in Europe Representative on Freedom of the Media explains:

“Female journalists and bloggers throughout the globe are being inundated with threats of murder, rape, physical violence and graphic imagery via email, commenting sections and across all social media.”

These threats undermine the safety and security of individuals who are active in the political sphere, in some cases causing them to live in fear, need to seek protection from the police or take extra safety precautions when in public. A recent example of this includes Laura Kuenssberg, political editor of BBC News, attending the Labour Party Conference with a bodyguard after an 18-
month campaign of abuse by online critics. Laura Boldrini, the Speaker of the Italian Chamber of Deputies, also outlined her fears for her daughter’s safety following threatening and aggressive posts in an interview with the Guardian highlighting again the impact abuse can have beyond the intended recipient.

Pushing people out of political spaces

As a result of online harassment and abuse, many individuals who are politically active online are choosing to self-censor, opt-out of participating in online spaces, or are being advised to shut down their accounts to prevent the abuse they receive. One recent example of this includes Owen Jones, British newspaper columnist and political commentator, who made the choice to quit social media after receiving threats of death and torture online.

Beyond online activity, abuse received on social media is also a major factor in politicians choosing to leave their positions or preventing people from entering political spaces in the first place. A 2017 survey conducted by BBC Radio 5 Live revealed that a third of all female MPs interviewed had considered quitting as a result of online abuse. The report by Atalanta also draws on research showing how women often cite the threat of online abuse and violence as a key factor deterring them from entering politics. Such impacts pose a serious threat to political representation in the UK when we consider that it is often women and minority groups that are most targeted.

Delegitimise leaders

According to the report by Atalanta, the intent behind online attacks is often to delegitimise politicians as leaders, especially in the case of gendered and sexualised attacks made against women. Even when this isn’t the intention the outcome is often the same. As stated by Dr. Mona Lena Krook in the same paper: “There’s a lot of research that shows that when you sexually objectify somebody, people tend to believe that the woman is less competent.”

For example, a study conducted by Funk and Coker in 2016 revealed that when a political candidate’s Facebook feed includes commentary on their body, potential voters are less likely to view the candidate as competent when compared to other candidates without such comments visible. This can have serious implications on the reputations of victims as well as democratic outcomes of elections.

Serve as a distraction

Another serious concern raised by the Atalanta report is “the fact that harassment and attacks on social media are a distraction from focusing on real, substantive issues that are central to the job of being a political leader.” Through the experiences of British Labour Party MP Jess Phillips, they explore how online abuse can distract politicians from their work emotionally and mentally, as well as in terms of the time victims’ often need to commit to filtering out comments or working with the police to locate perpetrators.

Explanations for rising online abuse

There have been many explanations of why online abuse of public figures has been increasing in recent years. This section briefly outlines some of those key explanations.

Perpetrators

Typically, perpetrators of online abuse have been understood to be psychologically or mentally unwell, or stereotyped as white, male “keyboard warriors” who are seeking attention and alleviating boredom. However, a recent study conducted by Ward and McLoughlin (2017) which analysed abusive posts made to MPs on Twitter suggests otherwise. Their research indicates that abusers actually have very close characteristics to wider society, and that perpetrators tend to be ordinary citizens who feel underrepresented by the current political system. For example, after conducting analysis on nearly 7,000 tweets, they found that the majority of abusive tweets were made in response to content posted by MPs and that abuse spiked around politically contentious discussions, such as Brexit. In addition, they also found that the majority of those posing abuse on Twitter were not serial transgressors.

The 6,952 abusive messages sent to MPs analysed in the study were sent by 4,775 members – which is an average of 1.4 abusive tweets per user. This, they argue, reveals that online abuse of MPs tends to be reactive, politically motivated and
driven by not having their voice heard. Not only does this render the stereotype of abusers as mentally unwell or attention seekers redundant, it also implies a failure in the British political system to fully represent the needs of people.62

Although this study offers us a useful insight into the perpetrators, the authors make clear that more research is required to identify abuser profiles and patterns of abuse. This has also been echoed by the police who have revealed that little investigation has been conducted into who the perpetrators are.63

Social media as a form of political engagement

Another explanation of why online abuse has been increasing in recent years lies with the use of social media as a tool for political engagement. Firstly, social media allows for a fast-paced, easy and emotive communication style which risks more direct and less thought-through statements being made.64 Secondly, many social media platforms allow for anonymity when making posts. This allows citizens to publicly express their opinions in ways that they may not usually be able to due to them being freed from the constraints of social norms and expectations.65

The British political sphere

A third explanation for rising online abuse of public figures can be found in the changing nature of the British political sphere. As discussed by Ward and McLoughlin (2017), there is increasing levels of polarisation and populism in British politics - both between and within political parties - which is in turn leading to more extreme political conversations and debate. This is fuelled by selective exposure (where a user can more easily access information online that suits their own pre-existing views, disregarding anything that challenges them), online echo-chambers (where “people of similar political outlooks spread or replicate each other’s messages but are rarely challenged by alternative viewpoints or voices”), and the anonymity allowed by many social media platforms.66

Reflection of broader social problems

A final consideration when addressing the causes of rising online abuse comes from understanding online abuse as a reflection of wider societal problems. As highlighted by Seyi Akiwowo, founder of Glitch, at a WebRoots Democracy seminar67, the online world and the offline world cannot be seen as separate spheres:

“The online world is a reflection of the offline world.”

This argument is evidenced by the fact that it is female MPs and MPs from minority groups that face the greatest amount of online abuse. This strongly indicates that any reforms put in place to address online abuse also need to be accompanied by efforts to understand and overcome growing prejudice and hatred in the UK.

A comment on data

It is important to note that much of the data drawn on in this chapter has come from Twitter. As highlighted by Josh Smith, from Demos, at our roundtable, this does not necessarily reflect that the majority of online abuse occurs on Twitter. Rather, it suggests that Twitter is more open to sharing its data than other social media platforms. Other social media platforms making their data available may help researchers and policymakers better understand who is being targeted and where, helping to inform strategies of prevention and protection.
policy roundtable

defining hate crimes online
responsibility
behaviour change
anonymity vs pseudonymity
As part of our research, we organised a roundtable at TechHub London for the *Kinder, Gentler Politics* report inviting a range of experts on online abuse including technologists, campaigners and journalists. The roundtable opened a discussion on the current state of online abuse, what has been done and what could be done by social media companies to tackle it. The list of attendees for this roundtable can be found in the appendix.68

**Defining hate crimes online**

The first part of the meeting was dedicated to defining hate crimes online. We asked how we should define hate speech and what the main issues are when defining it. What was highlighted very quickly by Alex Holmes from The Diana Award was that “it is not a definition, but a debate”, especially as there are nuances online such as GIFs and memes that aren’t being picked up on yet. The general consensus was that the definitions and laws already put in place defining hate speech and hate crime are proficient, but the enforcement of this is the issue.

Having a different definition for hate speech online from existing definitions offline was agreed to be problematic in tackling the issue of hate crime. An important point raised by Seyi from Glitch was that the debate “needs to move forward from theoretical discussion” as rape and death threats persist on social media and social media companies are “not doing enough” to tackle this immediate issue. As such, it was agreed that the focus on debating the definition of hate speech online is not conducive to tackling the issue and that existing definitions should apply in both online and offline spaces.

**Responsibility**

In the next part of the discussion we asked whose responsibility it is to tackle this issue. Specifically, we asked whether social media companies have the responsibility or the right to tackle online abuse when they aren’t based in the UK. From the discussions in the roundtable, it was clear that responsibility fell on three main groups; law enforcement, social media companies, and users themselves.

We discussed fines as a way of governments claiming responsibility for tackling online abuse. In 2017, the German parliament voted in favour of fining social media companies such as Facebook and YouTube up to 50 million euros if they fail to efficiently remove hateful posts soon after they are posted. It was agreed that this is a positive step forward. However, Chris Elliott from the Ethical Journalism Network highlighted that the legislation is “clunky” and as such social media companies crack down on many cases of free speech to avoiding paying fines. It also places the responsibility on social media companies to take down abusive content but leaves no incentive for users themselves to stop posting abusive or harmful content.

Simon James from the Cabinet Office’s Elections Team pointed out that there is a large amount of content that is harmful but not illegal and that it is harder to crack down on. Chris and Seyi added that taking down information from platforms is antithetical to a social media company’s business model. Therefore, another incentive is needed for social media companies to be more thorough and efficient in taking down harmful and abusive content. Other than social media companies, an incentive is also needed for users themselves to stop publishing and sharing abusive content and to instead report it.

This blurring line of responsibility becomes problematic when analysing nuances on social media platforms such as Twitter. Alex says this is the case with retweeting or sharing links to published content that is abusive. Are the users spreading this content just as complicit?

It was generally agreed that seeing but not reporting abuse fostered a “bystander culture” and that users witnessing abuse should also be responsible in reporting it. Attendees unanimously agreed that to be able to counteract online abuse efficiently, more needs to be done to prevent it from the source.
Behaviour change

The third section of the meeting discussed the steps needed to change behaviours surrounding abuse. All attendees agreed that “citizenship and media literacy are crucial” to helping tackle online abuse. Seyi suggested that the terms and conditions of social media platforms such as Facebook and Twitter should be made more accessible, through simple animations or by breaking them down into language easily understood by any 13-year-old. Furthermore, she believed that even using the word “rules” over “guidelines” would make users more responsible of what they publish.

Indeed, it was found that if institutions make their rules clear, people are more likely to follow them. This was discussed using the example of a study carried out on the Reddit community showing an improvement of online behaviour when policies were more visible. Rachel Fielden, from WebRoots Democracy, added that Reddit was a good example of a platform demonstrating digital citizenship. Each forum has a moderator regulating the discussion and users have the ability to down-vote content they find hateful, incorrect or against community guidelines putting the most down-voted content at the bottom of the forum.

Seyi said that ultimately, “we are all trolls on a spectrum because we’ve been conditioned to troll” which is primarily why having online abuse and safety integrated into citizenship teaching is important. There is not one specific character profile of a troll, rather we all have the potential to troll and publish hateful content and as such should all be taught how to use social platforms accordingly.

Anonymity vs. pseudonymity

We asked whether having an online identity would improve behaviour and make users more accountable for what they publish. To this, Seyi argued that people need to be protected and having no anonymity is “a very Western concept” that could not apply in countries where being able to retain one's anonymity is a matter of life and death. Samuel Rowe, from Yoti, believes in pseudonymity as a technical solution to online abuse, particularly when clamping down on creating new or fake accounts following suspensions from social media platforms. If the identity of an individual could be traced, the roundtable generally agreed that behaviour would change to make users more responsible for what they publish and share.

Both Chasidy and Seyi from Glitch saw that anonymity would hinder the possibility of empathy on the side of the victim and that it is being behind a screen that invites people to abuse others and not necessarily being anonymous.

As Josh Smith from the Centre for the Analysis of Social Media (Demos) pointed out, Twitter is so widely studied because of its open API. Therefore, comparatively, analysis on Facebook and Instagram is limited. In addition to this, Alex saw that Instagram would be harder to analyse for harmful content, whether they be suggested harms or harmful ideas relating to self-perception and body image; messages shared visually are not as easy to classify and remove.

Attendees all agreed that one of the key solutions to tackling online abuse is funding education on digital literacy and citizenship. First, in schools by integrating this education into the national curriculum and second, by making rules and procedure for online abuse and hate speech clear to all users.
consultation responses

stories and suggestions
As part of our research we held an open consultation asking for personal experiences and recommendations on tackling online abuse on social media platforms. We received over 100 responses from individuals across the political spectrum feeding into our study and final recommendations for politicians and social media companies to consider. The questions we asked gave us an insight into the kinds of abuse users received and how this impacted their willingness to participate in political discussions.

Alongside demographic information, respondents were asked questions about their confidence in voicing political opinions online, about whose responsibility they believe it is to tackle online abuse, and about their own experiences they may or may not have had. Just under half of respondents (49.5%) said they have been victims of online abuse in the past.

We had a varied range of respondents by age, gender, ethnicity, sexual orientation, and party affiliations. This breakdown is displayed in the following charts.

**Demographics**

**Gender**

**Age**

**Sexual orientation**

**Ethnicity**
Stories and suggestions

Suggestions from respondents who have not been subjected to online abuse

Of those who said they had not received abuse online before, 54% said they felt they had little to no confidence in voicing their political opinions online. The majority of these respondents said social media companies, in particular, have a responsibility to tackle online abuse. Many of them shared their own ideas and suggestions with us for how the problem could be tackled in future.

One respondent said there needs to be a focus on education as well as regulation, and that abuse is a “part of human interaction and can never truly be stifled.”

“Regarding minor abuse, that responsibility falls down to a combination of improvements in the social media platforms themselves to facilitate healthier conversations. My suggestion is some form of accessible education programme, or guidebook - teaching users how to break down and mitigate outrage/offense culture and how their own reactions contribute to polarisation. Also to teach people to never tweet/post in anger. Teach people to dig around news stories to spot fake news, and to not take information at face value. Teach people to also be tough to minor abuse or criticism - it’s a part of human interaction and can never truly be stifled. Citizens need to really develop the skills to use social media platforms in the most conducive way. Toxic abuse, should be regulated, and minor abuse should be educated about.”

Others argued that social media platforms should innovate to make users think twice about the content they are about to post online and to develop techniques to verify identities.

“When certain words are used in a post. There should be a small warning that you must accept before you press send. To make people think again before posting.”

“Authenticate users using a system like AirBnB.”

Some respondents expressed concern about the potential of state intervention and argued that the current generation is “oversensitive”.

“Centralised authorities shouldn’t have the power to silence speech. It’s too much of a slippery slope. Until the internet and social media platforms become more diverse and less centralised then they should also not be able to silence political views. Even if it is abusive. Should be regulated by the individuals who can request hate speech to be taken down.”

Experiences of respondents who have been subjected to online abuse

Despite being victims of online abuse in the past, this group of respondents reported marginally higher levels of confidence in voicing their political opinions online. 25% said they felt “very confident” in doing so.

The majority of abuse appears to have occurred primarily on Facebook and Twitter, however a number of other platforms were named including Reddit; 4Chan; Snapchat; Instagram; and Tumblr. Ask FM was singled out by one respondent as “truly the worst culprit when it comes to online abuse” stating that “the anonymity provided a platform for those who wished to abuse race, sexuality, gender etc.”

Many respondents shared personal stories and experiences with online abuse for inclusion in this report. One respondent said the comments had driven them to “suicidal despair”. Another was told by an abuser that they hoped she was “gang raped by the refugees that you are supporting” and another respondent who tweeted in support of LGBT equality was told to “kill themselves.”
“I liked a post about anti-racism and a family member trolled me, talking about how all Muslims are terrorists, that I’m an ISIS sympathiser and thanking God for Brexit as it will get rid of all the immigrants.”

“I was bullied by individuals who disagreed with my political opinions - deeply personal, hurtful insults, lies spread, hounded off certain bulletin boards etc. It pushed me into suicidal despair.”

“An ex and her new boyfriend used to leave abusive comments on my Facebook. Twitter is an endless cesspit of abuse and the London Young Labour group was previously full of abuse.”

“A leftwing activist contacted all local businesses to smear me, branding me a xenophobe, racist, atheist etc. I run a local forum in a leftwing borough (Lewisham) and it became known that I support Conservative politics and voted for Brexit. This led to a number of abusive reactions online from those who disagreed with my views.”

“I posted a tweet of solidarity for those at an anti-racism event that was picked up by some neo-nazis and I woke up the next day to about 90 notifications including rape threats, death threats and graphic images of decapitated bodies.”

“I was called a cunt a few times, been told to go fuck myself, and of course a feminazi!”

“I wrote an article published online and got abusive racist comments and tweets.”

“I would rather not go into specific detail but it was targeted at my disability and mental illness.”

“Some believe and have accused me of not caring about the vulnerable, working as a stooge for the establishment, that kind of stuff. Nothing overly abusive that I’ve felt intimidated however it is to such an extent that I worry young people will not wish to openly admit to being a Conservative or a Brexiteer which is a shame.”

Suggested improvements from victims

Those who told us that they used the reporting functions on social media platforms were asked about their experiences as well as their thoughts on what they think could or should be improved in future. Some expressed despair with one respondent stating that Twitter is “a toxic place” and that they “can’t imagine it ever improving.” Another said that their policies do not reflect “real life situations.” One respondent suggested that social media platforms could “prioritise minorities instead of cisgender, straight, white men.”

We also asked for general ideas, comments, or suggestions on what could be done differently by all relevant parties.

“Hate crime laws could be strengthened and protected characteristics broadened to include poverty, physical appearance, and weight.”

“It’s the responsibility of the social media companies to put in preventative measures that stop abuse from happening and to have a solid system in which abuse can be reported and taken down.”

“The government has a responsibility to educate people about the dangers of online abuse from a young age and to hold the social media companies accountable. Online abuse shuts down debates and contributes to mental health problems.”

“The tone set by media and the government enable and empower certain attitudes, especially when the government is ambivalent about its position on abuse.”

“The social media companies need better compliance policies to protect their users rather than the light touch approach currently in place. Accounts and users found to be abusive should be barred.”

“The constant creation of division and hatred from groups who look to benefit from it both in terms of campaigns, but also financially, is a huge problem. Groups like Westmonster (on one side) and The Canary (on the other) share stories devised to deepen rifts, using language and imagery to further develop the ‘them and us’ narrative. Having demonised the other side, they ask for donations to help fight them. Increasingly,
sources like this are the way people are experiencing the news. All they hear is negativity about the other side. About how evil and stupid they are. It’s not surprising that a daily drip fed diet of specific, manipulative and weaponised hatred leads people to abuse others online.”

“Social media companies have a moral obligation (if not legal too) to block online abuse using automated language analysis software. If an iPhone Siri app can understand my instructions then Facebook should be able to use lexical analysts to decode what people post for key words.”

Who should take responsibility?

Respondents were also asked where they believe responsibility for tackling the problem of online abuse lies. The most popular answers amongst both victims and non-victims of abuse were social media companies (30% vs 26%), abusers themselves (=18%), and the police (=17%).

The following chart compares the responses of victims (outer circle) and non-victims (inner circle).
focus groups

online bullying and intimidation
targeting of young women and ethnic minorities
negative impacts on political engagement
To investigate the matter of online intimidation and bullying more deeply, WebRoots Democracy and Royal Holloway University ran four focus groups with young people in London. We wanted to find out about their experiences of heated debates and online intimidation when discussing political issues, and to enquire if this affected their willingness to politically engage in various online platforms. Two of the groups were mixed, one was comprised of young women only, and the last was made up of young ethnic minority citizens.

Demographics

Gender

Sexual orientation

Online bullying and intimidation

The focus groups found that online bullying and intimidation were rife amongst young people. The vast majority of these groups of politically interested younger citizens had experienced some form of abuse when posting political views online and through social media. This was particularly true for young women and ethnic minority citizens, who appeared to be deliberately targeted by (largely anonymous) social media users. This latter finding is disturbing, though unfortunately not surprising, given the need to encourage greater political representation amongst these groups.

However, the participants drew a clear distinction between different types of social media. They were most commonly drawn to closed groups and online communities when expressing their
political views. In this regard, Twitter and Facebook were often avoided, in preference for WhatsApp, Instagram and Snapchat. One male (ethnic minority) participant reported:

“I’ve done it [talked about political issues on social media] but just wouldn’t do it again… never change anyone’s mind… can you imagine talking about knife crime on Facebook?”

A female participant summed up the general feeling: “I wouldn’t share anything [political] on Twitter or Facebook.” A male participant added: “Keep the more controversial stuff to your WhatsApp group.”

“Can you imagine talking about knife crime on Facebook?”

The often toxic atmosphere of open social media platforms was also linked to the problem of fake news, and the genuine fear that their views: “will be taken out of context, deliberately misconstrued… or dragged back from 5 years ago” [referring to the ‘footprint’ on social media that you are leaving behind].

Targeting of young women and ethnic minorities

Young women and ethnic minority citizens had been subjected to a wide range of abuse online and through social media when they expressed their views. Abuse often followed innocuous or positive postings by young people. One male, ethnic minority participant revealed that he had expressed support for #LondonUnited after the London Bridge terrorist attack and received a reply from an anonymous user, “it should’ve been you who died”. He said that he did not take the comment seriously, but it nevertheless illustrates the scale of the problem. As the same focus group participant, admitted: “Everyone gets trouble, but it tends to be people of colour and women.” – referencing the abuse dealt out to Diane Abbott.

“Everyone gets trouble, but it tends to be people of colour and women.”

Is the abuse a deliberately ploy to silence women and ethnic minority groups and diminish their voice in public conversations? One female, ethnic minority participant reflected that:

“A lot of the abuse is racialised… it’s trying to go back to the day when people of colour didn’t really have a voice… so when they abuse you, it’s definitely to try and get you out of that space they think you shouldn’t be in.”

Negative impacts on political engagement

Online abuse over political views seems to provoke two main types of reaction from young people. Either they retreat into closed groups or carefully self-censor the political content they publish on public platforms. One female participant stated:

“I express my views on social media, but I try not to do it on a public platform, because I don’t want to offend anyone… I wouldn’t like to go on the Internet to express my views… not that I have any outrageous views, but because I wouldn’t want to be attacked… because then you have to deal with ‘do you respond?’, ‘do you ignore?’”

This retreat (particularly amongst certain groups of young people) to closed groups, to get away from the poisonous political atmosphere that can pervade on open media platforms, is clearly dangerous for British democracy. In this scenario, the voices of young women and ethnic minority groups becomes less audible, while the voices of extreme groups and individuals are amplified. One male, ethnic minority participant told us:

“There was a debate about Palestine and I was, like, I don’t want to say anything and get smeared with the anti-Semitic brush…. I had to say very innocent things not to go down that route.”
This was a very politically engaged young man, who said that there are certain things he would not share online anymore.

A female participant emphasised how negative experiences had:

“made me very cautious... I don’t want to have to deal with all of that... it’s easy for me to sit here and discuss these issues with you all but I wouldn’t want to do it online.”

The young participants also underlined the toll that political activism online can take in terms of dealing with the abuse. Although there is a growing recognition of the stresses social media can create amongst young people, through bullying, intimidation and abusive behaviour more generally, the mental stress of online political engagement was emphasised by many young women.

“It had huge effects on my mental health for a long, long time.”

One female, ethnic minority participant, who was very politically active admitted:

“opposing these things [abuse and racist behaviour] – you need to have the emotional fortitude... and I sure as hell don’t... it had huge effects on my mental health for a long, long time, which I think is also not acknowledged a lot.”

Another young women told us:

“I’ve taken a detox... it causes so much anxiety and depression.”

There was a clear distinction between young men and young women in how they reacted to or dealt with the abuse. Perhaps this was because they received less abuse, perhaps it was because the abuse they received was less personalised, or perhaps it was because they naturally favoured a more combative approach. A number of male, white and ethnic minority, participants believed that “if you have political view you should stand by it” and not let the abuse affect you.

Ideas for change

A number of ideas for change sprang up from the conversation – about how things should change to improve the current situation. It was almost universally argued and agreed that social media companies should get a better grip on the situation and take action more quickly in blocking racist or offensive comments and the offending accounts.

Many expressed the view that social media companies should not regulate themselves, but should become accountable to similar regulation as is the case for TV and print media.

Kinder, Gentler Politics: Tackling online abuse in political debate
social media analysis

twitter influencers
facebook groups
existing and evolving policies
Given the number of existing and previous studies focusing on online abuse directed towards MPs, our report has instead looked at online abuse directed towards another key group in British politics, online influencers.

Influencers are primarily political journalists and campaigners and use platforms such as Facebook and Twitter to spread information, comments, and opinions on various hot topics of the day. Often, these influencers have a larger online following than MPs themselves and play a significant role in shaping political debate in the UK.

To provide an idea of the kinds of conversations happening online, and the kinds of abuse faced by online political influencers, we analysed the sentiment of more than 53,000 tweets directed at online influencers using both manual and automatic methods. For the automatic sentiment analysis the tweets we analysed covered a five-day period from the 31st of August to the 4th of September 2018. The dates for analysis were chosen at random as were the 60 influencers we looked at. Each of these influencers had at least 10,000 followers on Twitter.

Whilst automatic sentiment analysis is by no means perfect, it is the best and least resource-intensive method of analysing large quantities of tweets. However, this method of analysis cannot capture sarcasm, double negatives, slang, dialect variations, idioms, or ambiguity. It also cannot analyse abuse sent in the form of memes, GIFs, or videos.

Nonetheless, we undertook automatic sentiment analysis, using NVivo, of the tweets sent to our selected influencers in the run up to the 2017 General Election. Having delved into the detail behind the results, however, we deemed much of the output to be too inaccurate for inclusion or to have any significant meaning. A lot of the analysis was warped by the lack of context and recognition of who tweets were actually referring to. For example, on a closer inspection of the very negative tweets responding to Rachel Shabi, many of the abusive comments were in fact related to Labour MP, Margaret Hodge, who was being criticised for comments on anti-Semitism in the Labour Party. Some of the comments referring to Ms Hodge included her being a “nasty, vile, bitter hag” and a “poisonous troll”.

Looking at general, top line figures, the automatic sentiment analysis classified 64% of the tweets directed at influencers as “negative” or “moderately negative”. The remaining 36% was coded as “moderately positive” or “positive”. However, as stated, we considered this method of analysis to be imperfect and not rigorous enough to take meaningful, detailed, conclusions from.

We, therefore, aimed to understand sentiment better through manually analysing the replies to tweets from 47 of the 60 political influencers. To do this, we developed a scale ranging from a very enthusiastic, positive response to a threatening response, such as death and rape threats. Based on this scale, we analysed the replies of each tweet between the 23rd and 24th of July 2018. In total, we examined just under 10,000 tweets. Although the manual sentiment analysis allowed us to pick up on nuances, sarcasm, GIFs and videos that the software could not, the results of the analysis are still not generalisable. It would be wrong to use the results to claim that one user is more abused over another through analysing the tweets of one day.

Although, this method is also imperfect, the manual nature of the analysis allowed us to have
much greater confidence in the general, top line figures, providing an interesting snapshot into the kinds of online conversations political influencers are engaged in. Given the level of subjectivity involved and with more than one researcher working on the task, the data was generalised into three categories “positive interactions”, “neutral interactions”, and “negative interactions”. The analysis classified 10% of tweets as “positive interactions”, 66% as “neutral interactions”, and 24% as “negative interactions”.

Our researchers also noted a number of examples of the kinds of abuse that were being directed at political influencers.

Guardian journalist, Owen Jones, was subjected to homophobic abuse in response to one of his tweets:

“Owen, yes you’re gay, but have you ever thought that it’s a bit you know, odd to have a guy’s dick in your arse on a regular basis?” - @VHMOB

Kevin Maguire, from the Daily Mirror, received a threatening response to a tweet discussing the ISIS terrorist group known as “The Beatles”:

“Any chance someone could demonstrate the crimes they committed on you?” - @Takemineitsfree

Alastair Campbell, who was Tony Blair’s former Director of Communications, received numerous abusive tweets, including some which called him “a liar, murderer, and drunk”, “a treacherous drunk”, “a mass murderer” and another telling him to “keep your mouth shut.”

Facebook groups

Much of the discussion surrounding online abuse in existing literature has focused primarily on Twitter largely because of the open nature of the platform. Whilst conversations about politics take place all over the internet on social media platforms such as Mumsnet and Reddit, our report’s analysis is focused primarily on the two most prominent networks, Facebook and Twitter.

Discussions on Facebook can take place in public, closed, and secret settings. This may be in public spaces such as open groups or on the comment threads of pages. In more private settings, these conversations may take place in closed groups in which users have to request permission to join. In very private settings, these may take place in secret Facebook groups, which are inaccessible to researchers.

Facebook hosts a plethora of UK political discussion forums created and moderated by individual users. Often, these groups are centred on a political party, and occasionally use party branding, but are not run by officials of that party. Some groups have tens of thousands of members such as the “We support Jeremy Corbyn” forum (69,000 members) and the “British Politics” Facebook group (84,000 members).

For this report, our researchers joined a mixture of open and closed Facebook groups in order to gain a better understanding of the civility of discussions on the platform. Some groups were joined using our researchers’ own Facebook profiles, and others were joined using a fake account which was created in less than five minutes with no identity verification required. The account used a St George’s flag as its profile photo.

The five groups we examined were randomly selected and are a mixture of party-orientated and generic political discussion forums. The membership sizes range from 6,000 (Green Party Policy Discussion) to 85,000 (British Politics). The group sizes are displayed in the following chart.
Our researchers proactively sought instances of abusive content being published onto the forums. This was achieved by simply searching specific key words within the forums. The findings highlight some of the depths that political conversations online have sunk to. Across the different groups, we found examples of comments which were homophobic, racist, and sexist. A post calling for the assassination of London Mayor, Sadiq Khan, was identified in the UKIP Debating Society (7,000 members). This was published by one of the administrators of the group. The following quotes are examples of the types of abusive comments published within UK political Facebook forums.

**UKIP Debating Society (7,000 members)**

“Almost every negro I’ve spoken to sooner or later starts banging on about slavery such that one can tell it’s what drives them. Have whites not also been enslaved? Is there not still slavery going on in the world? But is it what drives us? No. Yet the slave trade, as taught in schools, is all about whitey enslaving black people.”

“Muslims should NOT be allowed to take up any position of responsibility, so we the BRITISH people can maintain OUR culture. Not the sort being imported into our country by the treacherous politicians. Boys and girls RAPEd by these EVIL people and their culture!!!!!!!!!!!”

“It is a fact that negroes have a different brain set up, created by the demands of living in Africa. It is nothing to do with racism to point out that after 8 million years of human evolution, those who stayed in Africa did not develop further than the stone age, as they were found 400 years ago. Although [Labour MP] Lammy was born in Britain, he was brought up by his Guyanese mother, after his father buggered off never to be seen again. He has done well getting to where he is, but I guess political correctness got him to where he is now.”

“Radical Muslim has been voted in by the overwhelming immigrant Muslim voters. We can only hope he does not live long enough to further the Islamic march to undermine traditional British culture. One can only pray for a lone wolf to do the country a great service. Just my own personal opinion and wish. ISIS supporter and creep to the Muslims who have systematically infiltrated London.”

“Once again London is lost to us, but as I have said before, its good in a way, to keep these traitorous MPs and Muzzies all in one barrel, so when the RAF decide enough is enough, they can bomb the shit out of the place!”

**UK Conservatives Online (11,000 members)**

“Kill all these bunch of terrorist Pakis.”

“I’m sending my TV back to Samsung…it’s full of left wing faggots!”

“Being gay is not natural but it is certainly fashionable. Mental illness is natural. Many people are mentally ill. Does not mean that mental illness should be celebrated.”

**British Politik (85,000 members)**

“What? You’re an expert because you’re a faggot? Please.”

“Left wingers are little faggots that rely on idiots like me to go to war after they invite problems in. I say we use liberals and democrats as OUR human shields.”
“Imagine Theresa May appointing this tranny misfit to her team! She’d be a laughing stock. Corbyn does it and all his sheep stand about clapping because they think it makes them even more colourful. Munroe ‘gay basher’ Bergdorf was appointed by Dawn Butler, the great big black mumma who sets a new record in having the lowest IQ in Parliament.”

**Green Party Policy Discussion (6,000 members)**

“Islamophobia is a stupid word. Islam is not a race it is an ideology. Stupid liberal cultural relativist nonsense.”

“It’s the Zionist owned press.”

**We Support Jeremy Corbyn (69,000 members)**

“It’s JUST a word, it originated from the fact that a lot of the black people from America – slaves predominantly – 2 or 3 hundred years ago were FROM Nigeria. Nigger is JUST a way of saying Nigerian in a slang sense! WTF is the problem with that???”

“You’re all riled now aren’t you? Point proven. It offends people and they can’t help but stick their oar in. THIS is the REAL issue of political correctness. Nigger. It’s nothing bad is it?”

“Snake in the grass. He gives pakis a bad name. There, I’ve said it.”

These examples represent a small fraction of the kind of content published to popular UK political discussion forums. Whilst it does not represent the majority of online discussion it highlights how low the quality of conversation can sink and how problematic the content can become. Many of these comments contravene Facebook’s own “Community Standards”, in particular with regards to its rules on hate speech.

Section 12 in Part 3 (Objectionable Content) of Facebook’s Community Standards states:

“We do not allow hate speech on Facebook because it creates an environment of intimidation and exclusion and in some cases may promote real-world violence.

We define hate speech as a direct attack on people based on what we call protected characteristics – race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity and serious disease or disability. We also provide some protections for immigration status. We define “attack” as violent or dehumanising speech, statements of inferiority, or calls for exclusion or segregation.”

Whilst group administrators are normally expected to act as moderators in the group, this is not always practical or effective, particularly with large groups or groups with numerous daily posts. As mentioned, one example we discovered in the UKIP Debating Society, an incitement to assassinate the London Mayor was made by the group administrator themselves.

Facebook groups have increasingly become the primary avenue for discussing politics online. The size and potential reach of these groups is also particularly significant. For context, the British Politics group has a membership size (85,000) is greater than the average daily circulation of the Manchester Evening News (37,000) and the Liverpool Echo (35,000) combined.

In September 2018, Labour MP, Lucy Powell, put forward a Private Members’ Bill sponsored by cross-party MPs including Nicky Morgan, Anna Soubry, Jacob Rees-Mogg from the Conservative Party and David Lammy, Ruth Smeeth, and Stella Creasy from Labour. The Bill called for social media platforms like Facebook to publish details of secret groups and ban more from being created.

Despite criticism from a number of technologists, the Bill passed unanimously to second reading. Explaining her case, Powell said:

“If 1,000 people met in a town hall inciting violence against a political opponent, or racism or hate, we would know about it and we would deal with it, our laws desperately need to catch up.

I’m proposing a small step establishing clear accountability in law for what’s published on online forums, forcing those who run the forums to no longer permit hate, disinformation and criminal activity.”

In a Spectator article, technology author and Director of the Centre for the Analysis of Social Media, Jamie Bartlett argued that the proposal would hit thousands of “harmless” groups run by “time-pressed hobbyists with zero knowledge of the law” and that “millions of us would be taking
on the legal responsibilities of a newspaper editor
with none of the time, money, staff or know-
how.”

Existing and evolving policies

Twitter

Of the major platforms, Twitter has arguably shifted the furthest in its attitude towards abuse and harassment. Throughout its early history, the company emphasised that its role in the political conversation was to provide a platform for free speech, declaring “we are the free speech wing of the free speech party”. The preamble to their original rules stated “we do not actively monitor user’s content and will not censor user content, except in limited circumstances.”

Around 2015 was when attitudes within the company shifted, with increasing media interest in the problem of harassment, and throughout 2016 and 2017, particularly, the site rolled out a number of initiatives aimed at addressing the issue. These included a new hateful content policy, enhanced mute and blocking options for individual users and enhanced reporting such as the ability to report multiple tweets at once. It strengthened rules of what constitutes abusive behaviour, requiring offending content to be removed, and permanent suspension for repeat violators. “Bystander reports” of abuse began to be treated with more significance, lessening the burden on the victim of abuse to address the problem.

Despite this, many users remain unhappy about Twitter’s inability to curb the amount of abuse on the platform. The site is criticised for not enacting further changes that have been suggested, such as the option for ‘can’t embed’ or ‘can’t quote’ tweets, reducing the chance of someone becoming a focus for abuse, or auto-blocking retweets and replies originating from an account that the user has blocked. Twitter’s curated ‘Moments’ feature has drawn criticism, as users are unable to opt out, and being featured has led to instances of being targeted.

In addition, some argue that Twitter faces an enforcement problem more acute than other platforms such as Facebook. The ease and anonymity of setting up an account means that large amounts of abusive content can be generated quickly, either by humans or bots. And whilst Facebook employs 7,500 dedicated content moderators on its site, Twitter has a total staff count of just over 3,000, of which only a small number are dedicated to policing content.

Facebook & Instagram

In recent years Facebook has significantly increased the resources allocated to content moderation; the job of vetting abusive, obscene or disturbing content. Staffing resources have significantly expanded, and Facebook is increasingly deploying machine learning to screen and remove content without the need for human moderators. In June, Facebook stated that as a result of these efforts, it had removed 2.5 million pieces of hate speech and disabled 583 million fake accounts globally in the first quarter of 2018.

However, whilst efforts to moderate content are being stepped up, controversy remains about what is classed as abuse, particularly in relation to public figures. Responding to a complaint from an MP in August 2018, Facebook stated “we do not apply our bullying policies to public figures because we want to allow discourse, which often includes critical discussion of people who are featured in the news or have a large public audience.” This has raised questions about the willingness of Facebook to clamp down further on abuse and harassment when it comes to public and political figures.

Facebook’s photo-sharing platform Instagram is currently the fastest-growing social network, though it is generally not regarded to host much political content. Earlier this year the site launched a new Bully Filter, applying Facebook’s DeepText tool to allow users to hide abusive comments.

Reddit

Reddit takes a more hands-off approach, due to the different structure of its platform. The site breaks down into thousands of individual communities (subreddits), each with its own set of voluntary moderators. Content is sorted by being voted up or down by other users. Reddit largely relies on these dynamics for subreddits to
effectively police themselves, only rarely making top-down interventions.

There are a central set of rules applying to all subreddits, and in 2015 after consultation with its users, Reddit announced measures to curb harassment, with the option to report offenders to site admins. In the announcement Reddit said it expected “no immediately noticeable impact on more than 99.99% of our users”.

Reddit doesn’t release information about its staff resources for content moderation, or the number of users who have been reported or banned. As Reddit accounts are easy and anonymous to set up, an outright ban would fail to present a major obstacle to anyone determined to harass.

Only on occasion has Reddit acted to take down and ban entire subreddits. In June 2015, five subreddits were taken down for harassing individuals. Although these subreddits were not overtly political in nature, it shows that the site is willing to act to enforce its rules. In the political realm, in November 2016, Reddit removed r/Pizzagate for doxxing (posting personal information), with the removal message “we don’t want witch hunts on our site”.

Kinder, Gentler Politics: Tackling online abuse in political debate
recommendations

civil internet tax
office for social media regulation
pseudonymous accounts
quarterly transparency reports
nationwide advertising campaign
investment in anti-discrimination campaigns
platform suspension powers
digital asbos
online abusers register
online neighbourhood watch
innovative t&cs
recommendations

Civil Internet Tax
Ofsoc (Office for Social Media Regulation)
Pseudonymous accounts
Quarterly transparency reports
Nationwide advertising campaign
Investment in anti-discrimination campaigns
Platform suspension powers
Digital ASBOs
Online abusers register
Online Neighbourhood Watch
Innovative T&Cs

During the design of our recommendations, we have focused on four key themes of i) resources, ii) oversight, iii) prevention, and iv) reprimands. Together, this comprehensive collection of radical proposals provides a strategy for society and social media to combat the rise of online abuse both in the short term and the long term.

Civil Internet Tax

What is it?

The Civil Internet Tax (CIT) would be a new tax levied on social media platforms to raise money to fund digital literacy and anti-discrimination initiatives as well as provide resources to help police forces address cases of online abuse. In addition, it would support the creation of a new regulatory body for social media platforms called Ofsoc (Office for Social Media Regulation).

Background

One of the fundamental stumbling blocks to a safer internet is the lack of available resources for digital literacy education, anti-discrimination campaigns, and policing. Whilst online abuse is primarily down to the behaviour of individual users, social media companies that financially profit from the content produced on their websites should shoulder some of the responsibility for tackling crime and abuse on their platforms. Online abuse should be viewed as a pollutant on the internet and should be treated as a negative externality. Political debate online should be conducted in a civil manner and every user should feel safe to express their political opinions free from the fear of hate crimes in retaliation.

This is a conclusion that has also been reached by the Digital, Culture, Media and Sport Select Committee who recommended the introduction of a levy on social media companies to finance digital literacy initiatives, a public awareness campaign, and extra resources for the Information Commissioner’s Office. The European Union have also been working on plans for a new tax on tech giants, nicknamed “the GAFA tax” (Google, Apple, Facebook, Amazon) which would be a 3% levy on global revenues. The UK Government has itself suggested taxing social media giants to help tackle extremism.

The underlying reasons for each of these suggestions has been that there are new strains on public institutions, caused largely by the rise of social media platforms, that require significant financial resources to tackle. Our CIT represents a levy that allows social media companies to contribute specifically to the significant public effort required to enable a more civil experience for users on their platforms.

With our proposal being based upon the number of users rather than revenues, it additionally recognises users as a product sold by social media companies to advertisers and is therefore be akin to a tariff. This method in many ways is a fairer tax than a tax simply on a company’s revenues which may generate significant funds from social media giants, but for most other firms would be a crippling tax. It may also act as an incentive for platforms to verify the identities of their users and ensure greater accuracy for figures provided to advertisers.

A similar idea has been proposed in the past by the Belgian economist, Paul De Grauwe, of the London School of Economics. His tax, however, is framed as a tax on ‘cost-free’ advertising revenue gained as a result of free content uploaded to the platform by users. For this, he suggested a levy of $10 per Facebook user which could be returned to the individual users each year, or for governments to use the revenue “to invest in education, the environment or
sustainable energy.” The CIT could be set at various rates, and could, if the Government went along with the EU proposals, work alongside a tax on global revenues.

Many of the proposals set out in this report will require significant funding and during a period of austerity, it is unsurprising that concerns about online abuse are not seen as a high priority for state institutions. We propose that funding for the ideas set out here come directly from the revenues raised from the CIT. At a rate of just £1 per user, the levy could raise up to an estimated £100 million for the UK Government from Facebook, Twitter, Snapchat, and Instagram alone.

**How would it work?**

In order to protect startups and smaller companies, the CIT should only apply to social media platforms which have at least a certain number of UK users, such as one million. The CIT would be levied based on the number of UK users and not on profit or revenue. The benefit of this system, compared to a tax on profits or revenue, is that it would make it difficult for tax avoidance initiatives to be used, whilst at the same time recognising the value of each individual user. We do not recommend a particular rate, however an example of the levels of tax revenue which could be raised is displayed in the table below, based on UK user estimates for major social media platforms.

<table>
<thead>
<tr>
<th>Company/ (No. of users)</th>
<th>£1.00 CIT</th>
<th>£1.25 CIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook (32m)</td>
<td>£32,000,000</td>
<td>£40,000,000</td>
</tr>
<tr>
<td>Twitter (20m)</td>
<td>£20,000,000</td>
<td>£25,000,000</td>
</tr>
<tr>
<td>Snapchat (13.6m)</td>
<td>£13,600,000</td>
<td>£17,000,000</td>
</tr>
<tr>
<td>Instagram (19m)</td>
<td>£19,000,000</td>
<td>£23,750,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>£84,600,000</strong></td>
<td><strong>£105,750,000</strong></td>
</tr>
</tbody>
</table>

For this to work effectively, it would be important to clearly define which platforms would be subject to the tax. This could be a role that Ofsoc (our proposed single regulator) could undertake in conjunction with HM Revenue and Customs. Considerations would include whether the platforms raise revenue or are not-for-profit platforms. Other considerations would be how to establish the number of UK users; for example, this could be achieved through self-reporting or using external audits.

**Ofsoc (Office for Social Media Regulation)**

**What is it?**

The Office for Social Media Regulation (Ofsoc) would be a regulatory body monitoring the practices of social media companies. The remit of Ofsoc may include overseeing how effective social media companies are in combating online abuse and to collate and analyse quarterly transparency reports. Ofsoc would also be a body that users could write to and raise complaints about abuse received on social media platforms, akin to the role of Ofcom and broadcasting complaints.

**Background**

Calls for a single regulator of internet companies have been growing in the UK. Currently, whilst laws exist that do apply to social media platforms, responsibility for oversight is often spread across various remits such as health, communications, and justice. In a green paper published by the think tank, Dot everyone, they make the case for an independent regulator to champion “new digital social contracts and uphold common standards across government, big tech, and emerging technologies.” In a letter published in the Telegraph, a number of broadcast and telecoms executives also called for the creation of a new social media watchdog to bring about “independent scrutiny of the decisions taken, and greater transparency.”

The interim report by the Digital, Culture, Media and Sport Select Committee notes that “there is no regulatory body that oversees social media platforms and written content including printed news content, online, as a whole.” Additionally, greater regulation over these platforms is something which Sharon White, Chief Executive of Ofcom has also called for. Regulators which
currently hold some responsibility over digital platforms include Ofcom (Office for Communications); ICO (Information Commissioner’s Office); CMA (Competition and Markets Authority); FCA (Financial Conduct Authority); and the Electoral Commission.

The creation of Ofsoc would help bring clarity over the rules which apply to social media platforms and act as a focal point of guidance for police forces, government departments, and members of the public. The new body could also assume responsibility for defining which websites would be classed as social media platforms and should be subjected to new regulations such as the Civil Internet Tax. The body could also review existing regulations on an annual basis and make recommendations to Parliament to ensure laws are kept up-to-date with technological advances. Amongst the suggested responsibilities set out by Doteveryone’s paper is the role of “developing mechanisms for consumers, citizens, creators, and civil society to seek redress for issues that fall in the gaps between existing regulatory mechanisms.” This is a reform which could help relieve the burden from both social media platforms and police forces across the UK.

If dissatisfied with the response of social media platforms to instances of online abuse, OFSOC could act as a next port of call prior to contact with the police.

**How would it work?**

Any new regulation of platforms will require oversight. With regards to our own recommendations, Ofsoc could be responsible for:

- Providing guidance to developers over pseudonymous accounts
- Administering the collation and analysis of quarterly transparency reports
- Initiating the nationwide advertising campaign
- Publishing guidance for law enforcement and members of the public about Digital ASBOs
- Overseeing the process for the anti-discrimination and digital literacy classes
- Managing the publicly accessible Online Abusers’ Register
- Publishing guidance and details for Online Neighbourhood Watch groups
- Advising the Government, Parliament, and other relevant institutions about social media suspension powers

Similar to Ofcom, Ofsoc could be subjected to oversight by parliamentary select committees such as the Digital, Culture, Media and Sport Select Committee, and have key appointments scrutinised and approved by Parliament. Partial funding for the new body could come from the Civil Internet Tax.

**Pseudonymous accounts**

**What is it?**

This proposal is for social media platforms to consider how best they can incentivise users to verify their identities and privilege those that do. The aim would be for social media users to have the option of being anonymous on the front end of the platform but for social media companies to know who their users are and whether they are real people as opposed to bots. This recommendation also calls on the Government to commission research into the links between anonymity and online abuse.

**Background**

The 2017 study into online harassment by the Pew Research Center found that anonymity is seen as a facilitating factor in encouraging the spread of harassment online. The American study found that half of those who have been harassed online had most recently been targeted by strangers or someone whose real identity they did not know. 89% of Americans said the ability to post online enables people to be cruel to each other.

Prince William, who spearheads a campaign against cyber-abuse, has also pointed the finger at online anonymity labelling it as “really dangerous”. A number of platforms such as Facebook and Google+ have introduced real name policies, but it is unclear how much impact this has had in reducing the prevalence of online abuse. Equally, many have highlighted the risk that removing front-end anonymity may have on certain vulnerable users. The use of government-issued ID cards to verify users’ identities is also controversial as it may leave behind certain users...
who do not have identity documents. In the same interview, Prince William states that “we are still responsible for our own actions online.” This points to another potential benefit of pseudonymity which is that even if verification does not reduce the likelihood of online abuse, it makes it easier for law enforcement and others to hold users accountable.

Other tech giants such as AirBnB, Bumble, and Coinbase have introduced identity verification for users, however those used by AirBnB and Coinbase require government-issued identity documents. Bumble’s system of verification, introduced to combat the phenomenon known as “cat-fishing” works by the user taking a selfie in a randomly select pose. The process is described by Bumble as follows:

1. Tap the verify button, either in your profile or on another person’s badge.
2. You’ll be prompted with an example of one of a hundred random photo poses by Bumble. Take a selfie mimicking that pose and send it to the Bumble administrators using the simple step-by-step verification process outlined in-app.
3. All photos are reviewed by a real person on Bumble’s team.
4. In a few minutes, you’ll receive a confirmation or rejection of your verification.
5. If a user has been reported as a potential fake profile and is rejected in verification, their profile will be turned off. This means they will see no one on Bumble and no one will see them.
6. If their photo is verified, the user will continue Bumbling as normal.

This is a potential method which could be adopted by social media platforms that would enable users to have an account, but not require a particular identity document. Using facial recognition software, it may also help the platforms block suspended users from creating a new account.

How would it work?

This two-part proposal is a loose recommendation for social media platforms to consider new methods in which they can verify all identities, or privilege those who do verify their identities, on their platforms. This is something that a number of social media companies are already looking into. It is also something that would be novel. For example, Facebook already operates a system of privilege for business pages that verify their accounts by uploading evidence, such as utility bills, regarding their companies. We would not advocate a system that makes it impossible for vulnerable groups from opening accounts on social media. Google also operates a similar system for companies. The second part of this proposal is a strong recommendation to undertake research into the link between anonymity and the prevalence of online abuse.

Quarterly transparency reports

What is it?

This would be a requirement for social media companies to submit quarterly transparency reports to Ofsoc including, amongst other things, detail on the progress made in tackling online abuse and hate speech on their platforms. They would also provide data on the number and types of reports of abuse received over the preceding quarter.

Background

One of the concerns raised within this study and elsewhere is the transparency of social media platforms regarding the levels of abuse reported and the counter-actions taken. In particular, the inconsistencies across platforms can create difficulties for researchers, journalists, and members of the public. Whilst some of the more prominent platforms such as Facebook and Twitter voluntarily produce transparency reports, these do not follow any set format and will therefore not always provide the information desired by interested stakeholders. In addition, there is no formal avenue for scrutinising the data and holding platforms to account.

This is an area in which the major players have a good story to tell. Facebook’s transparency report aims to answer the following four questions:

1. How prevalent are Community Standards violations on Facebook?
2. How much content do we take action on?
3. How much violating content do we find before users report it?
4. How quickly do we take action on violations?

Hate speech, defined by Facebook as “a direct attack on people based on protected characteristics - race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender, gender identity, and serious disability or disease” including “violent or dehumanising speech, statements of inferiority, or calls for exclusion or segregation”, is one of the main violations tracked. At the time of writing, there is no data available for the first question.

Twitter’s transparency report on the other hand, does not publish details on community standards violations or instances of hate speech but instead focuses on the number of legal requests for account information or content removal. Likewise, Reddit’s transparency report also only details legal requests for account information and content removal. This lack of information can make it difficult for regulators and others to understand the true scale of abuse on these platforms. A number of Amnesty International’s recommendations in their Toxic Twitter report were centred around better reporting of the nature and handling of online abuse, expressing frustration over the lack of clarity.

Mandatory transparency reports, set and scrutinised by Ofsoc, would help close this information gap and provide a clearer picture for public authorities as to the types and amount of potential hate crime taking place online.

How would it work?

Social media platforms, specified by Ofsoc, would be required to submit progress reports answering set questions every three months. To ensure that platforms participate, the regulator should be given the power to fine the companies that fail to respond.

Nationwide advertising campaign

What is it?

To help tackle the underlying issues behind online abuse, this proposal is for a nationwide anti-discrimination and digital literacy advertising campaign. The campaign would take place both online and offline. Akin to a public health campaign, the primary audience would be adult social media users in the UK.

Background

Campaigns for knife crime, cyberbullying, theft and even sexual harassment on the London Underground have been designed to inform the public of the issue and how to counteract it. To demonstrate that abuse committed online should be taken as seriously as abuse committed offline, the nationwide advertising campaign would act as an educative tool as well as a preventative measure against online abuse. If the Government is to be serious about stamping out abuse, it must recognise that those who commit abuse hold these views offline as well as online. Whilst social media platforms have a responsibility to ensure their forums are safe spaces, free from hate speech, the responsibility for tackling the underlying discrimination in society lies with the state. In lots of ways, this problem is easier to tackle amongst young people as they can be spoken to directly at school, colleges and universities. However, adult social media users who have left full-time education, and may have created accounts long after finishing education, are much more difficult to reach.

Whilst privileged white males can be subjected to online abuse, women and minorities can be subjected to abuse that is often more personal and more derogatory. We have included a number of graphic examples throughout this report. Despite much progress in combating bigoted practices in society, it is clear that many hold views that should be resigned to the previous century. To achieve zero-tolerance of such views, anti-discrimination campaigns should be an ongoing effort. The advertising campaign could help raise awareness of these issues, educate viewers, and kick-start an important national conversation about bigotry in society.

In addition, many adult social media users will not have had the opportunities afforded to young people today to become digitally literate citizens. A particular focus on digital literacy within these campaigns could also have a positive effect on online conversations, helping users understand
how to better discuss ideas online and how to handle instances of abuse.

**How would it work?**

The advertising campaign would take place offline on television, radio, public transport etc. as well as online on YouTube, Facebook, Twitter, and elsewhere. The adverts would be an open campaign against discrimination and point viewers to helpful resources. Based on the UK Government’s 2014-15 superfast broadband advertising campaign which took place online, on television, in print, and on radio, this proposal would cost an estimated £5m. Funds for the campaign could come directly from our proposed Civil Internet Tax.

**Investment in anti-discrimination campaigns**

**What is it?**

This proposal would earmark 25% of the annual revenues raised from the new Civil Internet Tax for investment into anti-discrimination initiatives in the UK. Similar to the nationwide advertising campaign, this recommendation represents a long-term investment in bringing communities together and combat some of the underlying root causes of abuse, both online and offline, that occur throughout society.

**Background**

Whichever approach current or future governments take, tackling online abuse will be a long-term endeavour. Some measures will be able to help reduce the risk to users in the short-term, but proper investment in permanently tackling the root causes of abuse will be required for the long-term. Whilst abuse can be directed at any member of society, some of the most vile abuse is often reserved for women and minorities. Proliferation of this kind of abuse online may have wide-ranging impacts in the long-term and could shape how individuals view each other in society. This recommendation recognises the long-term challenge and seeks to create an annual investment fund for anti-discrimination initiatives in the UK.

Under the 2010-15 Coalition Government, the Government Equalities Office (GEO) issued funding (from the European Commission’s PROGRESS scheme) of up to €250,000 to support anti-discrimination initiatives. Whilst this fund was primarily focused on policy ideas rather than grassroots campaigns, the objectives fit a similar mould to our proposal. The activities were particularly focused on the following topics:

- Non discrimination mainstreaming
- Diversity management in the public and private sector
- Roma
- Multiple discrimination
- Measurement of discrimination
- Discrimination on the grounds of religion or belief
- Discrimination of LGB and T people
- Communications and awareness training

One of the main objectives was to develop policies that “combat discrimination and promote equality beyond legislation.” Based on the estimate of £100m annual revenue from the CIT, our proposed investment fund would be a hundred times greater than the PROGRESS scheme described above and it would be issued on an annual basis.

In addition, the EU operates a fund under the Rights, Equality and Citizenship Programme to support anti-discrimination projects including those which “combat racism, xenophobia, homophobia, and other forms of intolerance.” The 2014-2020 budget for the programme is €439.5m, or €73.25m a year. Post-Brexit, it is unclear whether UK projects will have access to the programme or whether a new programme will be initiated in the UK. This proposal therefore could also solve a funding gap that may arise after the UK leaves the EU.

**How would it work?**

25% of the annual revenues raised from our proposed Civil Internet Tax would be earmarked to fund anti-discrimination projects in the UK. Based on the estimates set out earlier in the chapter, this could amount to £25m a year. The projects could be grassroot campaigns or policy reforms, but would be focused on tackling sexism, racism, homophobia, and other forms of discrimination in society. In particular, a focus...
could be placed on initiatives that combat discrimination amongst the adult population. As previously mentioned, this can be a more difficult to reach group of citizens when it comes to educational campaigns.

**Platform suspension powers**

**What is it?**

Platform suspension powers would act as a final sanction to be taken by the state to reprimand social media companies that consistently fail to significantly tackle hate speech on their platforms. This would be a radical reform requiring the input of Parliament, the judiciary, Ofsoc, and internet service providers (ISPs). The powers would block sanctioned social media websites from operating in the UK for up to three consecutive days.

**Background**

The internet is widely regarded as a borderless entity in which websites can operate freely anywhere in the world. Due to the nature of the internet, this is true to an extent, which makes it a difficult space to police effectively. It also weakens the impact of domestic laws. As things stand, social media companies (and others) have a free pass to operate their websites in the UK, with the state having little to no say in them doing so. This is not how it works for the majority of other companies that operate in the UK. For example, restaurants that ignore health and safety standards can be shut down on the recommendation of inspectors from local authorities. Another example is winding up orders which can be used by HMRC or other creditors. These powers can force companies into compulsory liquidation.

This final sanction of closing a company can help ensure compliance with laws and regulations. However, due to the size of the likes of Facebook and Twitter, and the fact that they are US based companies, it is difficult for the UK to impose any meaningful sanction that could drive significant reform. Following the 2018 Cambridge Analytica scandal, the UK Information Commissioner’s Office issued a maximum fine on Facebook for the misuse of data, amounting to £500,000. In the second quarter of 2018, Facebook’s global revenues were reported at $13.2bn, or around £10bn, making the fine equate to just 0.005% of the company’s global revenue.

Countries around the world which have introduced nationwide bans of Facebook in the past include Bangladesh; China; Egypt; India; Iran; Malaysia; Mauritius; North Korea; Pakistan; Sri Lanka; Syria; Tajikistan; and Vietnam. The fact that many of these countries contain autocratic governments and that decisions were often taken to close down political opposition, has rightfully cast a shadow over the potential of blocking social media websites. Sri Lanka introduced the ban earlier this year after racial hatred speeches being disseminated on the platform had caused riots across the country.\(^\text{112}\)

In London, last year, the ride-sharing tech giant, Uber, had their licence revoked by Transport for London (TfL) for failures in corporate responsibility.\(^\text{113}\) Whilst the decision was widely criticised in the media and amongst many users, it was a rare display of enforcement by a regional authority against a major Silicon Valley player. The ban was lifted once Uber made a number of changes to their policies on safety and security. Uber told the court that it “fully accepts that TfL’s decision was justified” and said the company has “since acknowledged and apologised for their past mistakes and made far-reaching changes to address them.”\(^\text{114}\)

For the state to effectively to protect citizens online, the powers to temporarily block social media websites in situations of endemic, unaddressed, hate speech on platforms could act as an important final sanction for regulators in the UK. It could also represent a new, more effective, incentive for social media companies to ensure their websites do not operate as a platform for hate speech. These new powers could also incentivise less notable social media platforms to engage with public institutions and regulators in a way that is not being done currently. Earlier this year, the then Culture Secretary, Matt Hancock, expressed frustration that only 4 out of 14 invited social media companies showed up to talks.\(^\text{115}\) In an interview with the Andrew Marr Show, he admitted that self-policing had not worked and that new powers were needed.
How would it work?

Many websites in the UK are already blocked by the state. These are often websites that host illegal content such as child pornography or breach copyright laws as is the case with a number of file-sharing sites. These websites are blocked using various methods via the “Big 5” UK ISPs. These are BT Group; EE; Sky Broadband; TalkTalk; and Virgin Media. Examples of websites blocked by court-orders include The Pirate Bay, Putlocker, and First Row Sports. The basis for these bans were that the websites hosted content that was illegal in nature.

For social media platforms, the platform suspension powers would be used as a very final sanction on companies specified by Ofsoc that have failed to comply with UK laws and persistently failed to tackle online abuse on their websites. The imposition of this power would need to be approved by courts in the UK with the decisions scrutinised by Parliament. The objective would not be censorship but to be a reprimand for platforms and a driver to engage with regulators in the UK. Our recommendation is that the maximum time period a suspension should last is 3 days.

Digital ASBOs

What is it?

Digital Anti-Social Behaviour Orders (DASBOs) would be a reincarnation of the ASBOs that were in place in the UK from 1998 to 2015, and still used in Scotland. The DASBO would operate in a similar manner to the Community Protection Notice (CPN) that replaced ASBOs in England, Wales, and Northern Ireland. The DASBO would prevent offenders from using certain social media platforms for a specific time period. In addition, those with DASBOs would potentially be placed onto a publicly accessible “online abusers register” (described later in this chapter). DASBOs would be preceded by a written warning and would be issued for less serious cases of online abuse.

Background

At the core of much of the debate surrounding online abuse, is the question of free speech and proportionate responses to violent language. The current status quo is that the vast majority of racism, sexism, homophobia and other forms of bigotry, goes unpunished. Some will argue that any type of sanction would be an attack on free speech and others will argue that inaction is to tolerate vile online behaviour. Our proposed DASBO would act more as a slap on the wrist, intended to temporarily ban users who engage in online abuse and encourage reform of their behaviour in future. Due to the scale of the problem, tackling hate crime can be an expensive task for the police and the judicial system. CPNs and ASBOs can be a cheaper alternative due to non-police bodies having the power to issue them such as local authorities and housing trusts.

The aim would be to prevent recurring behaviour of abuse and to protect others on social media networks. It would also be a more practical and active expression of zero tolerance towards hate speech in the UK.

This idea was previously suggested by the All-Party Parliamentary Group on Anti-Semitism in 2015 arguing that social media users who persistently spread racial hatred online should be given “internet ASBOs” blocking them from sites such as Facebook and Twitter. Their report drew a comparison with sex offenders stating that “there is an allowance in the law for banning or blocking individuals from certain aspects of internet communication in relation to sexual offences.” The report goes on to say that “if it can be proven in a detailed way that someone has made a considered and determined view to exploit various online networks to harm and perpetrate hate crimes against others then the accepted principles, rules and restrictions that are relevant to sex offences must surely apply.”

How would it work?

The power to issue DASBOs would lie with the police and Ofsoc. In the first instance the offender would be issued with a written warning, and should either the police or Ofsoc receive persistent complaints about an individual user, would a DASBO be issued. The offender would be required to follow certain rules such as deleting their social media account and ceasing use of social media websites for a certain time period. Failure to comply would result in the same punishments as a regular CPN (up to 2 years’
imprisonment or an unlimited fine for those over 18).

In addition, those who receive a DASBO may potentially be placed onto a publicly accessible “online abusers register” detailing their name and picture alongside exact quotations of abusive comments made.

Online Abusers Register

What is it?

The Online Abusers Register would be a publicly accessible register of individuals that have received a DASBO for persistently engaging in abusive behaviour online. Employers, in particular, would be encouraged to use the register during their recruitment process. Details of the user, the nature of the abuse, and the precise quote of abuse would be listed on the register. To encourage reform, an individual’s name would only be published on the register after a three month period and can be prevented if the offender attends classes on anti-discrimination and digital literacy. The aim of the register, therefore, would be to act as a disincentive to commit online abuse, protect citizens, and to encourage reform of the offender.

Background

The inspiration for this recommendation is drawn from speed awareness courses that can be taken by drivers in order to avoid receiving points on their licence. The recommendation provides an opportunity for the offender to reform and acts as a second chance. The aspiration, therefore, is not that there would be a long list of individuals with DASBOs, but to incentivise offenders to attend classes to help them reflect on their actions and reconsider the way they converse with others online in future. Just as the primary objective of the speed awareness courses is to encourage compliance with national speed limits, the objective of the online abusers register would be to encourage civil conversations online.

An evaluation of the speed awareness courses published in 2018 found that participation in the courses was more effective at reducing speed reoffending than the use of fines and penalty points. The study established that this was a causal effect of the courses rather than a result of differences between drivers who participated in the course compared with those who did not attend one.

A similar scheme related to low level crimes committed by drunk individuals has been introduced in Wales. The scheme enables people who have been taken into police custody after committing low level crimes whilst drunk to go on alcohol awareness courses instead of being taken to court and receiving a criminal record. Whilst the imposition of a DASBO would act as a sanction on the user, the register would act as a driver for reforming the individual in order to reduce the likelihood of reoffending.

How would it work?

Upon receipt of a DASBO, an offender would receive notice regarding the online abusers register as well as details of how they can book themselves onto anti-discrimination and digital literacy classes. These classes would need to be completed within a three-month period. Failure to do so would result in the individual’s name, photo, and offending quotation being published onto a publicly-accessible register of online abusers. Details would only be deleted from the register once the offender has completed the required classes. This process would be administered by Ofsoc.

Online Neighbourhood Watch

What is it?

The Online Neighbourhood Watch would be a scheme promoted by social media companies to incentivise users to report the abuse they see on the platform. The reporting of abuse would be encouraged through a rewards system integrated into the platform itself.

Background

Social media platforms already have policies on tackling online abuse and making the reporting process easier for users to access. However, this often places the burden of reporting the abuse on the victims themselves. Furthermore, the public nature of online abuse has been argued to create a toxic bystander effect that has absolved other
users of the responsibility of reporting abuse and breaks down a sense of community on social media platforms. The option to report abuse already exists but has not been incentivised fully to encourage users to also report the abuse they see.

The lack of community on social media platforms makes it less likely that people would report the abuse they see. Incentivising users to look out for one another by reporting the abuse they see would make users feel more protected and a part of a wider community. Furthermore, it would make moderating conversations easier for social media companies with the help of users to locate abusive content and remove it. In turn, this would make users feel more responsible for upholding a safe space and reporting users that publish abusive or harmful content.

As the name suggests, the inspiration for this proposal comes from the UK Neighbourhood Watch established in 1982. The scheme is a partnership that aims to bring people together to make their communities safer. It involves the police, local authorities, voluntary organisations, individuals, and families. The approach focuses on improved home security, greater vigilance, accurate reporting of incidents, and the fostering of a community spirit. A 2008 evaluation of 18 studies estimated that for every 100 crimes, an average of 26 crimes were prevented with the Neighbourhood Watch.¹¹³

The creation of an Online Neighbourhood Watch scheme would help social media companies foster greater vigilance and reporting by users. This could help reduce the burden on social media platforms expending financial resources in identifying instances of online abuse that should be removed under their own community standards.

How would it work?

Each social media platform would need to devise their own rewards system for this proposal to achieve maximum effectiveness. This would give users a status within the community for accurately reporting abuse they see, enabling a safer space for discussion online. Volunteers in the Online Neighbourhood Watch could be supported and guided by social media companies. In addition, they could be given privileges on the platform and be visibly verified as a trusted user. Whilst this is a relatively loose recommendation, it is important to note that communities of users who join together to report accounts do form, and this scheme would merely be an extension and a formalisation of what is already happening.

Innovative T&Cs

What is it?

This recommendation calls upon social media platforms to design new methods for gathering consent to the terms and conditions of their websites, ensuring that users understand the Community Standards they are signing up to.

Background

Tech giants are renown for disrupting the status quo and creating more effective methods of doing ordinary tasks. Despite this, the terms and conditions of using a social media website is not dissimilar to what has come before. Although the language and user experience is better than traditional services, it remains relatively hidden away and users are not forced to read them in detail upon creating an account.

A US study by Deloitte found that as many as 91% of users consent to terms of service without reading them, rising to 97% amongst 18 to 34 year olds. One study by York University, Toronto and the University of Connecticut even managed to get 98% of participants to agree to give up their first-born child as payment to join a fake social media network called Name Drop.¹²⁰

As mentioned earlier in this report, many of the major social media networks’ terms of service contain strong condemnation of online abuse and hate speech. However, these terms are often hidden away utilised more as legal protection for companies rather than a method of protection for users.

How would it work?

We do not propose any particular method but recommend that social media platforms invest in designing methods that gain informed consent for community standards and terms of service.
references
references

16. Online hate crime to be tackled by new national police hub, Home Secretary says, The Independent, October 2017.
17. Home Secretary announces new national online hate crime hub, Home Office, October 2017.
18. Countering illegal hate speech online #NoPlace4Hate, European Commission, July 2018.
21. Student who abused Fabrice Muamba on Twitter 'should not have been jailed', The Guardian, April 2012.
27. Germany votes for 50m euro social media fines, BBC News, June 2017.
29. We need more organisations to come out for trans equality, Stonewall, December 2017.
33. Naz Shah MP on the rise of online abuse in political debate, WebRoots Democracy, June 2018.
35. Online abuse of UK MPs in 2015 and 2017: Perpetrators, targets, and topics, University of Sheffield, April 2018.
36. Sexism, harassment and violence against women parliamentarians, Inter Parliamentary Union, October 2016.
38. Turds, Traitors, and Tossers: The abuse of UK MPs via Twitter, University of Salford, April 2017.
39. Analysis: Do ethnic minority MPs receive more Twitter abuse than their white counterparts? Evolution AI, April 2018.
40. Black and Asian women MPs abused more online, Amnesty International UK, accessed October 2018.
41. Jewish Labour MP posts anti-Semitic abuse she received online, The Times of Israel, April 2016.
42. Brutal stats show just how much more abuse ethnic minority politicians receive on Twitter than their white colleagues, Business Insider, April 2018.
44. Unsocial media: Tracking Twitter abuse against women MPs, Amnesty Global Insights, September 2017.
45. #HerNetHerRights: Mapping the state of online violence against women and girls in Europe, European Women’s Lobby, accessed October 2018.
68 The attendees at the “Kinder, Gentler Politics” roundtable were Areeq Chowdhury (WebRoots Democracy); Rachel Fielden (WebRoots Democracy); Maya Fryer (WebRoots Democracy); Khadija Said (WebRoots Democracy); Seyi Akiwowo (Glitch); Louisa Bartolo (The Diana Award); Chris Elliott (The Ethical Journalism Network); Alex Holmes (The Diana Award); Simon James (Cabinet Office); Chasidy Ochalek (Glitch); Samuel Rowe (Yoti); and Josh Smith (Centre for the Analysis of Social Media, Demos).

69 Why I am seeking to stamp out online echo chambers of hate, The Guardian, September 2018.

70 Lucy Powell’s bill is the wrong way to tackle online hate, The Spectator, September 2018.

71 Twitter’s Tony Wang: ‘We are the free speech wing of the free speech party’, The Guardian, March 2012.


73 Hateful conduct policy, Twitter, accessed October 2018.

74 How to use advanced muting options, Twitter, accessed October 2018.

75 Harassed on Twitter? You can now report up to five tweets at once, Wall Street Journal, April 2016.

76 Twitter today starts enforcing new rules around violence and hate, TechCrunch, December 2018.

77 MPs press social media firms over failure to take down hate speech, The Guardian, December 2017.

78 Twitter’s new features aren’t what users asked for, The Atlantic, September 2018.

79 Putting out the Twitter trashfire, Art Plus Marketing, February 2016.

80 Twitter’s new features aren’t what users asked for, The Atlantic, September 2018.

81 #TrollTracker: From tags to trolling, Atlantic Council’s Digital Forensic Research Lab, June 2018.

82 Hard Questions: Who reviews objectionable content on Facebook and is the company doing enough to support them? Facebook, July 2018.

83 “Did we create this monster?” How Twitter turned toxic, Fast Company, April 2018.

84 Hard Questions: Who reviews objectionable content on Facebook and is the company doing enough to support them? Facebook, July 2018.

85 Removing bad actors from Facebook, Facebook, June 2018.

86 Ibid.


88 Instagram now has 1 billion monthly users, now the fastest growing social network, Search Engine Journal, June 2018.

Promote ideas, protect people, Upvoted, May 2015.

Ibid.

Removing harassing subreddits, Reddit, accessed October 2018.


Disinformation and ‘fake news’, Digital, Culture, Media and Sport Select Committee, July 2018.


UK eyeing ‘extremism’ tax on social media giants, TechCrunch, January 2018.

Why Facebook should be taxed and how to do it, Social Europe, October 2017.


Regulating for Responsible Technology, DotEdvryone, May 2018.

Broadcasters want the UK Government to create an independent social media regulator, Gizmodo, September 2018.

It’s time to regulate social media sites that publish news, The Times, July 2018.


Bumble photo verification – Kiss catfish goodbye, Bumble, August 2017.

Facebook Transparency Report, Facebook, accessed October 2018.

Hate speech, Facebook Community Standards, accessed October 2018.


Reddit’s 2017 transparency report and suspect account findings, Reddit, accessed October 2018.

Apply now for funding to raise awareness about discrimination, Government Equalities Office, July 2013.


Sri Lanka blocked Facebook this week for allegedly spreading hate speech and violence, Gizmodo, October 2018.


Ten out of 14 of the biggest web firms snubbed government meeting to discuss rules on tackling online abuses, Mirror, May 2018.

Punish hate crime on social media with internet asbos, say MPs, The Guardian, February 2015.

Impact evaluation of the National Speed Awareness Course, Ipsos MORI, May 2018.

Drunk offenders will be able to go on alcohol awareness courses rather than be prosecuted, WalesOnline, June 2018.

Neighbourhood Watch, College of Policing, accessed October 2018.

You’re not alone, no one reads terms of service agreements, Business Insider, November 2017.

Kinder, Gentler Politics: Tackling online abuse in political debate
For enquiries, please email

hello@webrootsdemocracy.org