Tackling Online Abuse: Written evidence submitted by Antisemitism Policy Trust on 15/09/2021 (TOA0008)

For more than ten years, the Antisemitism Policy Trust has been seeking to inform and improve the structures and facilities for addressing online abuse, including antisemitic and other forms of hate against those with legally protected and non-legally protected characteristics.

This submission sets out a number of reflections and recommendations for consideration by the Petitions Commission on online abuse. It includes an evaluation of the scale and impact of online abuse on internet users, the Government's Draft Online Safety Bill and the legal and technological solutions to address online abuse. The Trust liaised with key partners in developing the response, including the Community Security Trust.

The Antisemitism Policy Trust is grateful to the Committee for its consideration of online abuse and strongly supports the introduction of a new regulatory framework to deal with online harm and any action to tackle the impact it has on users.

The scale and impact of online abuse on internet users

Online harm is a part of the continuum of violence, as set out by criminologist L Kelly. It can manifest as verbal, physical or sexual abuses of power,[1] with online harms constituting part the verbal part of the spectrum (at least in the first instance). It can be threatening, harassment, intimidating, name calling and threats for violence. The importance of online abuse on victims of minority groups cannot therefore be understated. Whether antisemitic, gendered or intersectional abuse, cyber hate has long term and devastating impacts on its victims. As academics have identified, incidents of abuse online can have a lasting effect, left in public to be seen by a victim's peers and

[1] Ray and Smith, 2001, 'Racist offenders and the politics of hate crime, *Law and Critique,* Vol.12, pp.203-221

others unknown to the victim, often increasing the harm impact.[2] Empirically, Professor Paul Iganski, in analysing victim's experiences of crime, found that 36% of those targeted with a bias motivated crime, such as hate crime, found that overcoming the incident was "very difficult" compared to 13% of those targeted in a non-bias incident.[3]

The number of antisemitic incidents occurring online has grown exponentially. The Community Security Trust (CST) recorded that in 2015 there were 185 such incidents, and 697 cases in 2019,[4] an increase of 277%. Average annual quotas for online antisemitic incidents sit at around 40% of the total number recorded. These incidents have been reported into CST and either victim or perpetrator must be based in the UK to be counted. If CST were to trawl for antisemitism it would have too many reports to count. Antisemitism online ranges from overt antisemitism, to the often legal, but equally harmful, antisemitic tropes. Antisemitism Policy Trust research has revealed that negative stereotypes for Jewish people searched on Google include "why are Jews so greedy" and violent searches, which count for some 10% of the 170,000 antisemitic searches each year, include "Jews must die".[5]

The ongoing Coronavirus pandemic has highlighted the way the internet is used to abuse minority groups online. Antisemitic content manifested online has included conspiracy theories about the origins of the disease, blaming Jews for a fake conspiracy or using the virus for nefarious purposes, to outright calls to infect and kill Jews.[6]

As part of the Antisemitism Policy Trust's efforts to tackle antisemitic abuse online, we have explored its nature and impact. Specifically, our work has highlighted that gendered antisemitism is widespread. The case for exploring intersectional abuse online is evident. Of more than 9,000 threads on the neo-Nazi web forum Stormfront about feminism, 60% mentioned Jews.[7] Similarly, on alternative 'free speech' platform 4Chan, Media Matters for America – an NGO with which we partnered - found a huge increase in overlap between antisemitism and misogynist posts.[8] Karla Mantilla, in her

---

[2] Danielle Keats Citron, *Hate Crimes in Cyber Space*, 2014
[3] P Iganski, 'Hate Crimes Hurt More', *American Behavioural Scientist,* Vol. 45, No.4, pp. 626-638
[4] https://cst.org.uk/data/file/9/0/IncidentsReport2019.1580815723.pdf
[5] https://www.antisemitism.org.uk/wp-content/uploads/2019/02/APT-Google-Report-2019.1547210385.pdf
[6] https://antisemitism.org.uk/wp-content/uploads/2020/05/Coronaviru-antisemitism-May-2020.pdf
[7] https://www.antisemitism.org.uk/wp-content/uploads/2019/05/5982-Misogyny-and-Antisemitism-Briefing-April-2019-v1.pdf

book ‘Gendertrolling: How Misogyny went Viral’, highlights that ‘gendered trolling’ has three dimensions: it targets women who express opinions, it includes graphic sexualised and gendered slurs, and it occurs at a high level of regularity and intensity.[9] Further research from the United Nations Commission in 2015 found that 73% of women surveyed had experienced online abuse, with 18% of those who had experienced abuse calling it serious internet violence.[10] Amnesty in a study of online abuse against women worryingly found that 36% of women who had experiences online abuse or harassment felt their physical safety was threatened; highlighting the vast impact online abuse has on its victims.[11] Working To Halt Online Abuse (WHOA), a United States based group, found that of the 3,393 individuals who reported cyber harassment between 2000 and 2011, an overwhelming 72.5% defined as female, compared to 22.5% as male.[12]

As hate crime expert Joanna Perry argues, the victim-centric approach to hate crime is key, but oversimplifying the victim experience by focusing on one strand of hate, as opposed to examining intersectional hate, risks undermining our understanding of the harm caused.[13] This equally applies to the response to online hate and abuse. Cyberspace should therefore not be exempt from offline-space standards[14] and the same rules against bullying, harassment and hatred against minority groups must be applied or extended.

Government proposals to tackle the issue

Since the committee began its deliberations, the Government has published the draft Online Safety Bill. Though the legislation is promising and marks an important step in the fight against online harms, there are areas of the Bill we believe require attention in order for it to be fully effective.

Duties Of Care

---

[8] https://www.antisemitism.org.uk/wp-content/uploads/2019/05/5982-Misogyny-and-Antisemitism-Briefing-April-2019-v1.pdf
[9] https://antisemitism.org.uk/wp-content/uploads/2020/03/Web-Misogyny-2020.pdf
[10] https://antisemitism.org.uk/wp-content/uploads/2020/03/Web-Misogyny-2020.pdf
[11] https://antisemitism.org.uk/wp-content/uploads/2020/03/Web-Misogyny-2020.pdf
[12] Danielle Keats Citron, *Hate Crimes in Cyber Space*, 2014
[13] Joanna Perry, ‘At the intersection: hate crime policy and practice in England and Wales’, *Safer Communities,* Vol.8, No.4, 2009
[14] Danielle Keats Citron, *Hate Crimes in Cyber Space*, 2014

We strongly support the proposals to establish a 'duty of care' set out in the draft Bill. The Government's initial documents (the Green and White papers) proposed that such a duty would encompass regulated bodies taking "reasonable steps to keep their users safe and tackle illegal and harmful activity on their services" and referenced "reasonable and proportionate action" to tackle harms on their services. The duties as they stand are more specific, and do not reference risks of harm that are "reasonably foreseeable". This, in our view, is an error. An all-encompassing duty, applied differentially according to size or other key factors, would likely provide more protection and ensure a systems-facing approach, rather than the perspective on content which the current drafting risks.

So far as the proposed new duties are concerned, they apply differently to companies according to 'size and functionality' which in turn will determine an organisations placement in Category 1 (which has specific considerations in respect of legal but harmful content) and Category 2 (which does not). The draft Bill omits Search Services from category 1 duties. This means that irrespective of size, a search company will not be bound by the duty to address legal but harmful content, nor to protect content of democratic importance, for example. The only duties that Search Services will have will be in relation to limiting or removing illegal content from their platforms.

Research by the Antisemitism Policy Trust and Community Security Trust has demonstrated, on more than one occasion, that Google's services have led to harm.[15] Fo r example, before campaigners forced a change, Google prompted users towards the search 'Are Jews…. evil?' (whereby 'evil' was autosuggested by Google), and its 'SafeSearch' feature cannot remove antisemitic images. Microsoft Bing, meanwhile, sent users to the search term 'Jews are b******s' before being alerted to the problem. Amazon's Alexa service directed people to both antisemitic and anti-Muslim results before it was embarrassed into changing its systems.[16]

[15] https://antisemitism.org.uk/wp-content/uploads/2020/06/APT-Google-Report-2019.1547210385.pdf
[15] https://antisemitism.org.uk/wp-content/uploads/2021/05/Unsafe-Search-Report.pdf
[16] https://www.dailymail.co.uk/news/article-8991925/Amazon-says-investigate-claims-Alexa-gives-racist-answers-questions-Jews.html

The ‘size and functionality’ test set out above may well leave many alternative platforms out of category 1, even if they host large volumes of harmful material. Platforms including Bitchute, Gab and 4Chan, house extreme racist, misogynist, homophobe and other extremist content that radicalises and incites harm. The Community Security Trust has outlined in detail some of the most shocking and violent materials on these sites and whilst illegal material has been present, much of that content is legal but harmful (and would be addressed in other environments, such as a football ground, cinema or on TV/radio). That lawful material can and has transferred to more mainstream platforms and has influenced real world events. The relevant schedule (4) needs amending, for example, to reference the risk register developed and maintained by Ofcom in Clauses 61 and 62.

Even in its current form, we do not have confidence that the duty placed upon category 1 platforms will be fully effective. There are no minimum standards set out for the Terms and Conditions, expected to have “dealt with” harm to adults. In the past, these have proven to be hugely inconsistent across platforms. Terms and Conditions for addressing harmful content should be required to meet a minimum standard and the wording of the Bill should be amended to recognise this, including by defining what ‘dealt with’ means. Furthermore, risk assessments of harmful content performed by companies in scope should meet a requirement to be ‘reasonable’ to prevent gaming of the system.

There is an explicit exemption from all duties for content present on the website of a recognised news publisher. This is deeply problematic. The Antisemitism Policy Trust has worked with Government and others for many years, to highlight the abuse on newspaper website comment forums. For example, as secretariat to the APPG, the Trust worked with the Department of Communities and Local Government (now MHCLG) towards a guide delivered by the Society of Editors in 2014[17] which was inspired by discussion of this form of harm. The exemption in Clause 18 relating to newspaper comments boards should be removed or, at worst, amended to ensure publications have measures in place to address harm on relevant boards.

[17] https://www.societyofeditors.org/wp-content/uploads/2018/10/SOE-Moderation-Guide.pdf

Penalties

Senior management liability is included in the draft Bill, but only as a reserved and fairly limited power relating to specific information retrieval. Fines are important but will be written off as the cost of doing business by major companies. Senior management liability, as we have in financial services, can be a powerful and effective tool that will encourage companies to comply with the law and face consequences. It is therefore our position that Ofcom should be granted full enforcement powers with associated criminal sanctions under Chapter 6 of the draft Bill in relation to senior management liability, for breaches of the duties of care, in extremis.

Clause 96 details provisions for publishing decisions by Ofcom but there is no provision to mandate publication of a breach notice by a service. The Antisemitism Policy Trust believes publications, like newspapers directed by IPSO or others, should have details of their breaches of a duty of care available to view on the platform.

Legal and technological solutions to take action against people who commit online abuse

Several pieces of UK legislation can be used to prosecute illegal antisemitic online harms. However, many of these pieces of legislation passed prior to the development of social media and fall short of what is required to protect the public. Several cases of online hate have been successfully prosecuted in the UK, but these form only a fraction of the total cases recorded by third party reporting services.

Cases prosecuted in the United Kingdom include:

- The conviction of John Nimmo, who targeted former MP Luciana Berger with death threats online, including that she would "get it like Jo Cox". Nimmo was found guilty of nine offences under the Malicious Communications Act 1988 and sentenced to two years and three months imprisonment.[18]
- Alison Chabloz uploaded antisemitic songs on YouTube, including videos mocking Holocaust survivors. She also posted videos which called the Holocaust a "bunch of lies" and labelled the Auschwitz extermination camp a "theme park". Chabloz was found guilty of three offences under the Communications Act. She was sentenced to 20 weeks suspended sentence and banned from social media for two years.[19]
- Following a history of internet trolling and targeting Jewish social media users, such as former Member of Parliament Luciana Berger, far-right activist Joshua Bonehill-Paine was found guilty of publishing written material intended to stir up racial hatred under the Public Order Act. He had posted online grossly offensive images such as a negative caricature of a Jewish man next to Auschwitz with a bottle of "Roundup" weed killer spraying him, as an advertisement for "an anti-Jewification event". Other images included a poster calling to "Liberate Stamford Hill", an area with a high proportion of visibly orthodox Jews. Bonehill-Paine was sentenced to 3 years and 4 months imprisonment.[20]

---

[18] https://www.bbc.co.uk/news/uk-england-tyne-39008963
[19] https://www.wiggin.co.uk/insight/high-court-rules-on-conviction-for-offences-under-the-communications-act-2003-in-relation-to-the-posting-of-a-hyperlink-to-and-a-video-of-offensive-material/
[20] https://antisemitism.org.uk/wp-content/uploads/2020/03/web-online-harms-briefing-2020.pdf

Legislation in the United Kingdom

The law should be an expression of society's commitment to pluralism, tolerance, equality and, crucially, human dignity. Punishing hate, and therefore online abuse based on protected characteristics, is crucial to upholding these principles.[21] The importance of perceived victimhood, by the victim or any other person, as set out in the MacPherson Inquiry following the murder of Stephen Lawrence, is central to bias motivated crime, hate crime recording and related operational concepts in the United Kingdom, such as investigation. This should allow for intersectional victim centred legal approaches and policy. Additionally, the serious nature of online hate crime, as outlined in section one of this submission, and abuse online more widely, should take harm caused to the victim into consideration. Regulation of online spaces has been argued to restrict freedom of expression and speech,[22] but online platforms are privately owned, and already place their own restrictions on speech. It is not correct to suggest they are entirely free speech spaces. Furthermore, expressions of hate are already limited in the offline sphere, and supress some seeking to engage in open discourse, restricting speech in other ways for minority voices.

The Antisemitism Policy Trust is not calling for a change in criminal law to tackle antisemitic hate crime but has provided detailed evidence to the Law Commission which is reviewing Hate Crime laws. However, criminologists and legal experts have raised concerns about the implementation of current legislation to tackle all forms of hate crime. This includes concerns about addressing online abuse. Judges and juries are often left to decipher what 'hostility', the term used in UK law for hate crime, means to determine an offender's guilt.[23] Guidelines from the Crown Prosecution Service, sentencing guidelines for judges and juries, training for police in evidence collection and victim liaison and training for other agencies must be recommended and put into practice in order for any legislation to tackle online abuse and hate to be workable and effective in protecting victims. CPS guidelines relating to legal guidance for racial and religious hatred do not mention the online sphere.[24] The do however specify general the

[21] Walters, 'Conceptualising Hostility for Hate Crime Law: Minding the Minutiae when Interpreting Section 28(1)(a) of the Crime and Disorder Act 1998', *Oxford Journal of Legal Studies,* Vol.34, No.1 (2014), pp.47-74
[22] Danielle Keats Citron, *Hate Crimes in Cyber Space*, 2014
[23] Walters, 'Conceptualising Hostility for Hate Crime Law: Minding the Minutiae when Interpreting Section 28(1)(a) of the Crime and Disorder Act 1998', *Oxford Journal of Legal Studies,* Vol.34, No.1 (2014), pp.47-74

use of "derogatory language towards ethnicity, race, nationality or religion, (including caste, converts and those of no faith)".[25] These need to be expanded to reference and focus on online and social media-based hostility.

Legally, there must be definitions of online abuse offences against protected characteristics in order to adequately protect the victim. Racial and religious abuse online can be considered both a substantive offence (it is illegal in its own right) and a general penalty enhancement office (any criminal offence can be seen as aggravated by racial and religious factors). Internationally, in terms of the Organisation for Security and Coordination in Europe (OSCE) framework, hate speech offences are not included in their hate crime concept. However, nationally, substantively, racial and religious abuse is an offence according to the Racial and Religious Hatred Act, 2006. Section 28 of the Crime and Disorder Act defines the meaning of "racially aggravated" meaning, any offence as outlined in Sections 29-32 can be uplifted due to racial or religious hostility. Offenses include assault, criminal damage, public order offences and harassment. In terms of sentencing provisions, Section 153 of the Powers of Criminal Courts (Sentencing) Act 2000 requires courts to consider racial and religious hostility as an aggravating actor for any offence. This would include crimes online such as those included in the Malicious Communications Act (amended 2001) and the Protection from Harassment Act, 1997.[26] However, for other forms of online abuse, such as disability, gender identity and sexual orientation no substantive law applies. A crime has to be committed and proved in court, such as online harassment or public order offences, which can then be uplifted by aggravated factors as stipulated in the Criminal Justice Act 2003. This means that penalty enhancements are only applied at sentencing and are not necessarily addressed at the evidentiary stage. Hostility might be investigated and considered during the prosecution stages because of the perception-based recording of the crime earlier in the criminal justice process. However, the awareness of disability hate crime and gender identity hate crime may mean this is not considered, as these forms of hatred are not specifically a hate crime offence in and of themselves. This can mean that those who commit online abuse are not always being charged, prosecuted and convicted for hostility motivated crimes.

---

[24] https://www.cps.gov.uk/legal-guidance/racist-and-religious-hate-crime-prosecution-guidance
[25] Ibid
[26] https://antisemitism.org.uk/wp-content/uploads/2020/06/web-extended-online-harms-briefing-2020.pdf

Technological Solutions

Technologically, we cannot expect victims to simply 'turn off their computer' and turn away from the abuse. Legal and technological solutions must put the needs of the victims at the forefront of the solution. When the computer is switched off, the comments or abuse is still there for the victims' peers and wider communities to see.[27] Technological reform cannot pin it's hope purely on those committing online abuse.[28] With the ability for perpetrators to shield their true identities, whether with pseudonyms, the use of Virtual Private Networks (VPNs) or other Internet Protocol (IP) anonymising tools, technological remedies for victims and those impacted by online abuse must be examined.

As distributors of information, it is important for online internet sites to incur liability as a distributor of information. It should not matter if a person reporting and objecting to abusive content is harmed themselves.[29] Hateful speech should be met with a consequence for a commercial platform, as Facebook and others have learnt to some cost through the Stop Funding Hate campaign. This action, to date, has almost always been taken by non-governmental actors, with no formal legal liability.

Technological solutions which require automated and machine learning efforts to combat online abuse are problematic. Former Facebook Content Policy team head, Dave Willner commented that "automation is better with spam than with issues about what something means".[30] Therefore the requirement for technology companies to increase their capacity for human review is crucial to curb both legal and illegal online abuse. The training of those human moderators is crucial and should be quality controlled.

---

[27] Danielle Keats Citron, *Hate Crimes in Cyber Space*, 2014
[28] Ibid
[29] Danielle Keats Citron, *Hate Crimes in Cyber Space*, 2014
[30] ibid

Platforms such as Wikipedia have introduced novel recruitment of users, as opposed to human reviewers overseen by the social media companies, to enforce community norms. Wikipedia's user-led compliance model allows new users, once they have shown their compliance with the community standards, rules and other regulations of the platforms, can apply to become administrators with layers of power to create locks to stop other actors on the platform engaging in wrongful editing.[31] This could be explored in relation to social media based platforms.

Anonymity online has arguably allowed the spread of online abuse from unaccountable users, bots and other malicious actors. Chris Wolfe, an American attorney who specialises in internet law said, on anonymity; it is a useful tool to help speakers avoid retaliation but not when it is solely used for destructive speech.[32] The Antisemitism Policy Trust has made recommendations that it should be the responsibility of a given platform to determine the degree of anonymity it wishes to offer users, though there should be a risk-assessed approach, with incentives against hateful content, and severe penalties for anonymous abuse.

In our view, if a crime or a libel has been committed in the UK on what will in the future be regulated technologies, and companies in scope cannot or will not provide proof of identity, where a magistrate's court order demands it (subject to an appropriate burden of proof), then a range of options should be considered. The Trust believes that the civil or criminal liability should pass to the platform itself (this would be in line with existing measures in the e-Commerce Directive), and fines or other corrective measures could be put in place. We would suggest giving the platforms a year to become compliant. Companies should apply the 'Know Your Client/Customer' principle, familiar to those in the financial sector, with appropriate safeguards. Using some of the legal framework required by companies offline, such as Customer Due Diligence in The Money Laundering, Terrorist Financing and Transfer of Funds regulations 2017, online companies should verify users' identities before allowing use of their platform. This should be done even if use of the platform is free of charge and when users are not

[31] ibid
[32] ibid

regarded as ‘customers’. Middleware technology could be employed to protect users true identities until such time as illegal behaviour occurs.

An alternative is that social media companies could employ algorithms to prioritise non-anonymous activity in users’ feeds. Non-anonymous actors can be assumed to have more accountability.[33] This technique is used on some message boards, and theoretically Facebook requires users to use their real identities, but this is not always the case. This would allow users who have not made any infractions, whether anonymous or not, to remain on the platform but would encourage them to de-anonymise their accounts. Anonymity is therefore a privilege that can be lost[34] when an actor engages in hateful abuse online. We know that it is crucial to allow anonymity online as some users, such as domestic abuse victims, refugees and those in vulnerable situations will use the platform.

Studies have investigated the use of site design alteration on limiting harassment online, and the same could be applied for online abuse. For example, if platforms could nudge a user to not post hateful content, or create a longer posting process, by creating a cooling off period between submitting a post and it being published on the platform it could allow a user,[35] engaging in online abuse due to spur of the moment emotional motivations, to cool off, edit the post or remove it before the damage is done. This would be particularly advantageous for younger social media users. We know, for example, that Public.io is currently investigating taxonomies of harm for the Government, and hope that the companies work will provide some helpful information in this regard.

[33] ibid
[34] ibid
[35] Danielle Keats Citron, *Hate Crimes in Cyber Space*, 2014