



Item 5: Towards a safer space for minorities: positive initiatives to address online hate speech: the role of national human rights institutions, human rights organizations, civil society and other stakeholders

Dr Andre Oboler, CEO, Online Hate Prevention Institute

Email: ceo@ohpi.org.au Web: <https://ohpi.org.au> Facebook: <https://fb.com/onlinehate> Twitter: [@onlinehate](https://twitter.com/onlinehate)

In considering how we address online hate, one thing is clear: wishful thinking is not enough. Neither are memes and videos opposing hate, positive initiatives that bring people and communities together, education programs for school, public advocacy or public education. Counter speech is not enough. Empowering young people is not enough. These are the dominant approaches of recent years, they are what we have promoted even as the hate continued to rise.

If we want to get on top of this problem, we need take it more seriously. We need to invest both economically and through political capital in real solutions. We've seen major shifts in recent years in areas such as privacy. We need a similar shift when it comes to tackling online hate.

I've worked at the cutting edge of tackling hate speech in social media since a small platform called Facebook, then with a mere 70 million users, was struggling to compete with myspace. The journey highlights many different positive initiatives to tackle online hate speech. It also highlights some of the roles national human rights institutions, human rights organizations, civil society and other stakeholders, including governments and platforms, can play.

In 2008 Israel convened the Global Forum for Combating Antisemitism, a gathering of civil society experts, academic experts, politicians, and community leaders from around the world. It included a panel on Internet Antisemitism where I released my work on Antisemitism 2.0 and the normalisation of hate through social media.¹

¹ Andre Oboler, 2008. "Online antisemitism 2.0. 'Social antisemitism' on the 'Social Web'," *Jerusalem Center for Public Affairs*, number 67. Online: <https://jcpa.org/article/online-antisemitism-2-0-social-antisemitism-on-the-social-web/>

Lesson 1: Governments can convene forums and give a platform to civil society.

When the Global Forum met again it included a range of working groups. I was honoured to co-chair the working group on Internet and Media antisemitism with David Matas from Canada from 2009 until 2018. Each time we met with global experts we compiled a report. Each report listed key incidents, key areas of concern, positive initiatives, proposals for action, and open challenges.² The contents was compiled ahead of the meeting from the participants, with recommendations then discussed and refined when we met.

Lesson 2: With the right structure, a broad range of experts and secretarial support, a solid picture can be created even as technology and threats rapidly evolve.

A key challenge listed in 2009 was the lack of metrics or even an approach to measuring antisemitism in social media. In 2011 a proposal on the units of measurement, and design documents for software to crowd source data and its classification was brought to a the Working Group and stakeholders from civil society and government across multiple countries provided feedback and discussion.

Lesson 3: Defining grand challenges spurs work on solutions.

Lesson 4: The solutions are strongest when they can be discussed with peers and build on feedback.

In 2012 we took the project focused on online antisemitism and turn it into a standalone civil society organisation, the Online Hate Prevention Institute, with a remit to tackle all forms of online hate. There was ground break work in 2012 on “Aboriginal Memes and Online Hate”,³ in 2013 on “Islamophobia on the Internet”,⁴ work on antisemitism identified blind spots in platform policies.⁵

Lesson 5: Much of what we learn tackling one form of online hate is equally applicable to other forms of online hate. Supporting all minorities makes methodologies and tools more effective more quickly.

Lesson 6: It takes deep technical expertise to do this work, particularly to keep up with new platforms and technological changes. There are a real skills gap. We can't expect every minority community to fill the gap, we need dedicated civil society organisations with the appropriate skills and a broad perspective working for all.

Lesson 7: We need collaboration between specialist civil society organisations and the people who live the reality of the online hate and the organisations that represent them.

² <https://mfa.gov.il/MFA/AboutTheMinistry/Conferences-Seminars/GFCA2013/Pages/Online-Antisemitism-A-systematic-review.aspx>

³ Andre Oboler, *Aboriginal Memes and Online Hate* (Online Hate Prevention Institute, 2012). Online: <https://ohpi.org.au/aboriginal-memes-and-online-hate/>

⁴ Andre Oboler, *Islamophobia on the Internet: The growth of online hate targeting Muslims* (Online Hate Prevention Institute, 2013). Online: <https://ohpi.org.au/islamophobia-on-the-internet-the-growth-of-online-hate-targeting-muslims/>

⁵ Andre Oboler, *Recognizing Hate Speech: Antisemitism on Facebook* (Online Hate Prevention Institute, 2013). Online: <https://ohpi.org.au/recognizing-hate-speech-antisemitism-on-facebook/>

The software was used for the Global Forum in 2015, producing a draft report of empirical data on antisemitism across Facebook, YouTube and Twitter.⁶ Importantly, all the data was all saved creating a valuable archive of the antisemitism of that time. It was also reported to the platforms in the usual way. We monitored the sample of over 2000 items for 10 months to see what came down and when. We found complete lack of consistency within platforms and between platforms. At the 9-month mark we offered the platforms the data in return for them reviewing the sample again. Two of the three platforms took us up on the offer, but little came down. The platform was later entirely rewritten to enable it to be offered a service, configurable for different types of hate, running in different languages and able to be embedded on organisations own website while still storing data in a central repository.

Lesson 8: User reporting that is external to the platforms is essential, it is the only way to share data on what people have reported between experts.

Lesson 9: To engage in tackling online hate, local civil society organisations need access to sophisticated reporting systems. They also need staff with the time and expertise to engage on the topic of online hate. Even when access to the software is offered for free, engagement remains a challenge. Funding to monitor online hate is scarce, and often tied to developing new unproven systems which often fail to meet expectations. The expertise needed to work in this space is underappreciated and difficult to appropriately fund.

Lesson 10: Data needs to be kept. This allows peer review, archives, and future research.

Lesson 11: Content needs to be reported in the usual way so the effectiveness of platforms responses can be tested. This enables real and independent transparency reporting.

Lesson 12: Trusted flagger systems can stifle civil society if failure to conform with platforms views of what breaches policy may result in a loss of the privileged status.

The Online Hate Prevention Institute began focusing on briefings to cover more ground more quickly. A briefing is an article that looks at a particular problem, for example hate against a particular minority on a particular platform. It typically gives 5 to 8 examples and explains the problem and why the content should be removed. Each item has an screenshot, but also a link so it can be loaded live as long as it remains online. We encourage visitors to use the link to load the content and report it. Briefings often included either step by step reporting instructions, or a link to reporting guides.

Lesson 13: Short focused work on specific issues can have as much impact as long reports.

Lesson 14: Giving minority communities content to report, and instructions to help them do so, increases their resilience and empowers them for the future.

Lesson 15: Publicly archiving platform failures and explaining the gaps in their policies, with examples, helps them close the gaps and holds them publicly accountable. The media plays an important role in creating that accountability. Visible watermarks branding content as hate speech examples help to prevent recirculation.

⁶ Andre Oboler, *Measuring the Hate: The State of Antisemitism in Social Media* (Online Hate Prevention Institute, 2016). Online: <https://mfa.gov.il/MFA/ForeignPolicy/AntiSemitism/Pages/Measuring-the-Hate-Antisemitism-in-Social-Media.aspx>

Lesson 16: Redacting the name and photos of people who post the hate prevents a backlash that may be disproportionate. Others take a different view and believe in naming and shaping as a deterrent.

Large scale reports are still important to provide a wholistic view of major issues and recommendations for action. One recent report looked at four terrorist attacks in 2019 which were linked to 8chan: Christchurch, Poway, El Paso and Halle.⁷ Our work in briefings and rapid reports exposed the methodology, analysed the extremist messages, and led to work actively calling on those hosting copies of terrorist manifestos to remove them. The public may assume this work was done by government, but a small NGO on the other side of the world can play a major role in advancing public safety. The report has many recommendations for different kinds of stakeholders. One which stands out is the need for a system where platforms can hand over concerns about a user to a relevant government agencies for treatment according to law if the deterrence ability of platforms is insufficient to prevent further harm.

Lesson 17: The threat to minorities from online hate can extend all the way to terrorism.

Lesson 18: Countering online hate is by nature both local and global work. NHRIs, civil society groups, academics, politicians, and technology companies can all help make urgent work a success.

Lesson 19: Extreme harm or repeated abuse by platform users should see platforms transfer matters to relevant government agencies. System are needed to enable this.

We've also engaged in doctrinal legal research, examining how the law should approach the question of online hate.⁸ Some ideas are that the law should aim for 'online / offline consistency'. There is a 'principle of generality', that well-established laws should also apply online. The 'principle of inclusion' holds that conduct that is unlawful offline should not escape criminality simply for being online. The 'principle of appropriate adaptation' holds that conduct permitted offline should only be criminalised online if the fact of it being online changes its nature or prevalence.

Lesson 20: As governments increasingly look towards regulate as a solution to online hate against minorities. Laws tackling hate speech should have a solid doctrinal basis.

We've engaged in sociological research, investigating the types of contributions different stakeholder types are best placed to make and how different stakeholders can support each other.⁹ There is transfer of burden from community members to civil society organisations when reporting takes place and this can be harmful when community organisation are powerless to take action. A systemic approach, supported by technology, can convert this to a system that builds resilience and supports action.

Lesson 21: Many different types of organisations have a role to play in tackling online hate, but the best results come from structured collaborations that created effective synergies.

⁷ Andre Oboler, William Allington and Patrick Scolyer-Gray, *Hate and Violent Extremism from an Online Sub-Culture: The Yom Kippur Terrorist Attack in Halle, Germany* (Online Hate Prevention Institute, 2019). Online: <https://ohpi.org.au/hate-and-violent-extremism-from-an-online-subculture-the-yom-kippur-terrorist-attack-in-halle-germany/>

⁸ Andre Oboler, 'Legal Doctrines Applied to Online Hate Speech' (2014) 87 *Computer and Law*. Online: <http://www.austlii.edu.au/au/journals/ANZCompuLawJl/2014/4.pdf>

⁹ Andre Oboler and Karen Connelly, 2018. "Building SMARTER Communities of Resistance and Solidarity", *Cosmopolitan Civil Societies: an Interdisciplinary Journal*, 10(2), 91-110. Online: <http://dx.doi.org/10.5130/ccs.v10i2.6035>

We've looked at Artificial Intelligence, for our own use, as deployed by social media platforms, and as used by civil society to research and measure online hate.¹⁰ One thing is clear, human creativity will always find away to defeat the technology. While automated solution can significantly reduce the volume of hate, it will take the constant vigilance of experts to defeat the creative efforts of people dedicated to promoting hate.

Lesson 22: The dialogue between platforms and civil society needs to be ongoing. Civil society organisations play a vital role in monitor changes in online hate and alerting platforms to new manifestations of hate. The nature of hate varies with the targeted minority and the country where the hate originates. There is a need for a large number of local partnerships. Some platforms, like Facebook, are starting to do this. Others are yet to reach this point.

We unfortunately regularly find ourselves and those we work with under attack. Sometimes it is from hate promotes who seek to undermine our work. Other times it is from algorithms that see us use all the same names and words as extremists and are unable to recognise the context. Social media platforms are certainly profiting from online hate, I know this simply from the amount the Online Hate Prevention Institute had to pay to ensure our content addressing this problem gets seen. This is perverse. Specialist civil society organisations working on online hate are not only acting as janitors for these mega corporations, we are paying them for the privilege.

Lesson 23: Verified civil society organisations working to tackle hate and extremism need to be “allow listed” to provide context to algorithms and time poor content moderation staff.

Lesson 24: Those working to tackle online hate need better protection, support and recognition for their role on the front lines tackling online hate.

Lesson 25: Social media companies, either directly or through government intervention, should fund the work of civil society to tackle problems on their platform.

Lesson 26: Work combating hate speech should not be regarded as a “social issue” and should be except from geographic based promotional restrictions on platforms like Facebook. The defence of universal human rights across borders is vital to ensure political and/or ideological use of hate speech can be called out by those who are safely able to do so.

There are many other aspects of our work. Public talks, training activities, conference engagements, supporting universities with academic research, supporting community organisations under threat, serving on government delegations, supporting policy discussion, providing submissions to consultations and inquires (government / department / parliamentary / NHRI inquiries / etc). We have also designed plans to support large numbers of civil society organisations with young people we would train, but who would be embedded in a minority community organisation. We have also developed regional plans to support local teams of civil society and academia working across many different kinds of hate speech and across many languages but with shared methodologies and tools. We have also developed a global

¹⁰ Andre Oboler, “Technology and regulation must work in concert to combat hate speech online”, *The Conversation*, 12 March 2018. Online: <https://theconversation.com/technology-and-regulation-must-work-in-concert-to-combat-hate-speech-online-93072>

plan for tackling online antisemitism with international partnerships.¹¹ Crisis also occur and demand immediate responses, like the rise of online hate during COVID.¹² Many working in the space burnout in the fact of constant demand and minimal support.

Lesson 27: There is a shortage of deep civil society expertise in online hate and a huge demand. We need plans and resources to increase civil society's capacity.

Lesson 28: We are not short of plans, but we have a gap between our expertise and our resources. We need to fund not only research, but action. That action needs to go beyond counter speech, beyond education and beyond empowering young people. Those are all important but they are not sufficient.

On the topic of fake news, misinformation and disinformation, hate is often based on ignorance and ignorance is compounded with misinformation and insinuation. The history of antisemitism highlights this, from Blood Libels to the Protocols of the Elders of Zion. The seeds of many modern conspiracy theories have their roots in antisemitism, and their target remains the Jewish community. Holocaust denial can be traced right back to the Nazis. It was Himmler, head of the SS, who on April 14th 1945 ordered that "surrender is out of the question... no prisoner should fall into the hands of the enemies alive". Holocaust denial is not just about the past, it is about the future. The message of the deniers is that "the Holocaust didn't happen, the Jews deserved it, and we need to finish the job." It is a clear threat. I've been raising the problem of Holocaust denial in social media since 2008. Only last year did YouTube ban it, only in recent months did Facebook ban it, but worst of all, Twitter claimed to ban it and only for their CEO to then claim it was still permitted. Recent work by the American Jewish Congress have highlighted the gaps in Twitter's policy and examples of the content that continues to circulate. From memes to Holocaust denial propaganda videos.

Lesson 29: Governments are increasingly adopting the IHRA Working Definition of Antisemitism, it is rapidly becoming a de facto standard. Those who have not adopted it should do so.

Lesson 30: Social media companies should clearly and explicitly ban both antisemitism and Holocaust denial. They should commit to using the IHRA Working Definition of Antisemitism and the IHRA Working Definition of Holocaust Denial and Distortion as tools to assist them in recognising and responding to online antisemitism.

Similar definitions and broad consensus are needed on the hate affecting many other minorities. The work on antisemitism provides a model that should be applied to protect other communities as well.

Lesson 31: IHRA itself has recently adopted a Working Definition of anti-Roma hate, again something for Governments and Social Media Platforms to adopt and put into practise.

¹¹ Andre Oboler, "Solving antisemitic hate speech in social media through a global approach to local action", *An End to Antisemitism!* Vienna, Austria, February 18-22, 2018. Online: https://anendtoantisemitism.univie.ac.at/fileadmin/user_upload/p_anendtoantisemitism/Articles/Article_INT_Oboler.pdf

¹² Andre Oboler, David Wishart and Mark Civitella, *Online Hate Prevention Institute Supplemental Submission: Online Vilification and Coronavirus* (Online Hate Prevention Institute, 2020). Online: https://parliament.vic.gov.au/images/stories/committees/isic-LA/inquiry_into_Anti-Vilification_Protections/_Submissions/Supplementary_submissions/038_2020.06.17_-_Online_Hate_Prevention_Institute_Redacted.pdf

There is much more work to be done to protect many more minorities, but a trail has been broken and there is a path ahead of us.

My final point is on the role of religious leaders and faith actors. The Beirut Declaration on Faith for Rights highlights how faith can support human rights. Faith leaders can play an important role in reducing the tensions between communities and build respect for all people. They can also go further, bringing communities together to tackle online hate. The Council of Christians and Jewish (Victoria) has worked with the Online Hate Prevention Institute over the past year to run a training program teaching people to recognise and tackle many forms of online hate against minorities. Hate 2.0 has normalised hate within online spaces and the community, beyond the online world it is in our local communities where we must reverse this trend. Faith communities, inspired by their leaders, can make a real difference in reversing that trend.