



## **HATE SPEECH IN CANADA: A NEW DEMOCRATIC THREAT REQUIRING POLICY INCENTIVES**

**Date:** May 16, 2019

*Disclaimer: This briefing note contains the encapsulation of views presented throughout the evening and does not exclusively represent the views of the speaker, or the Canadian Association for Security and Intelligence Studies.*

### **KEY EVENTS**

On May 16, 2019, the Canadian Association for Security and Intelligence Studies (CASIS) Vancouver hosted Dr. Heidi Tworek at its roundtable meeting titled “Hate Speech in Canada: A New Democratic Threat Requiring Policy Incentives.” Dr. Tworek is an Assistant Professor of International History at the University of British Columbia. She is also a non-resident Fellow at the German Marshall Fund of the United States, the Canadian Global Affairs Institute, and the Joint Centre for History and Economics at Harvard University. She works on the history of news and of international organizations. Alongside academic publications, she also writes about German and transatlantic politics and media for a wide variety of venues including *Foreign Affairs* and *Wired* magazine.

In her presentation, Dr. Tworek highlights the increased prevalence of hate speech in Canada, particularly on social media platforms, where both legal restraints and ineffectual corporate prevention strategies are exacerbating the problem. She later cited notable international examples aimed toward the effective reduction of hate speech, as well as holding social media corporations responsible for content regulation on their respective platforms. The roundtable discussion that followed examined the consequences of amending Canadian Charter rights to better address hate speech and the possibility that a myopic legal focus on hate speech may inspire individuals to join right-wing extremist (RWE) groups.

## NATURE OF DISCUSSION

### Presentation

Dr. Tworek's presentation began by distinguishing the differences between hate and harmful speech, and cited examples of the latter. There was an emphasis on social media's dominant role in propagating both hate and harmful speech and stated that hate speech on social media was a threat to democratic integrity. The presentation concluded with describing and evaluating the various efforts to combat hate speech in Europe and speculated on how Canada may implement similar measures in the future.

### Question & Answer Period

The question and answer period yielded discussion on the following: redefining identitarianism, coordinated attacks carried out by VTSMs, whether social media is changing how we engage in democracy, echo chambers online, and specific inquiries of law enforcement techniques.

### Roundtable Discussion

The roundtable discussion centered on the general need and desire to amend Canadian Charter rights to combat the prevalence and impunity of how hate speech is dispersed. It also explored the propensity of RWE groups to embrace a narrative of victimization to attempt to justify extremist rhetoric. Anecdotal examples of directly challenging methods of disinformation were provided and well-received.

## BACKGROUND

### Presentation

Dr. Tworek opened by comparing present hate speech trends to the Nazi's use of propaganda as information warfare that allowed them to rise to power. She states that the Canadian government has revisited its consideration on if it should regulate hate speech through social media and is now looking at how to do so. Dr Tworek offers a definition of harmful speech as:

“online communication that is abusive, threatening, denigrating, or that incites violence, and which may therefore undermine peoples' full, free and fair participation in politics and political debates.”

Harmful speech can take a variety of forms. Two specific forms were given particular attention: doxing, which is the search for and publishing of private or identifying information (about a particular individual) on the Internet, typically with malicious intent; and swatting, which is the action or practice of making a prank call to emergency services in an attempt to bring about the dispatch of a large number of armed police officers to a particular address. Of prime importance, Dr. Tworek notes that harmful speech has the potential to target minorities and expose them to vicious attacks and silencing online resulting in one-dimensional discourse and propagation of disinformation.

According to Dr. Tworek, the executives who own social media platforms, like Facebook and Twitter, are becoming increasingly content agnostic: they are refusing to definitively punish behaviour that would injure the advertising revenue stream. In doing so, these platforms harness the psychological motivations behind ego, hatred, and fear by enticing the user to stay online. Dr. Tworek quoted statistics which showed that one in four Canadians was said to have been the target of hate or harmful speech online, and a 600% increase in reported hate speech has been noted in Canada between 2015 and 2018.

A key problem in learning how to combat online hate speech is that social media platforms are refusing to disclose data to researchers, leading to confusing conclusions. Dr. Tworek makes the example that Twitter, at a glance, has a disproportionately high amount of hate speech in comparison to any other social media platform. This is because Twitter uses a more powerful application program interface (API) to analyze and identify its content.

Dr. Tworek proposes that harmful speech is a threat to democratic integrity, because small groups of motivated actors can apply a disproportionate amount of pressure to larger groups, shouting them into silence. Harmful speech accomplishes this by creating barriers for participation, promoting conflict, tension, and distrust, and by threatening and silencing at both the individual and group level. In so doing, these acts tear the fabric of democratic communication and the principles of equitable and free discourse online.

The legislative solutions available within Canada fall under four legal families: criminal, civil litigation, human rights, and election law, which all possess common flaws in prosecuting hate speech online. Chief among these flaws are the lack of enforcement and cost effectiveness of proposed measures. Further, the definitions of hate speech online are too narrow and punitive for the enforcement of mass infractions. Jurisdiction poses a problem as well, with

virtual private networks (VPN) and traditional state borders, police cannot typically apprehend, let alone prosecute perpetrators of hate speech overseas.

When Facebook and Twitter were founded, the platforms' approach to content moderation was ad hoc and performed almost entirely on a case-by-case basis with no clearly defined metric of what content was inappropriate for the platform. To date, most of these "takedowns" are performed by AI, guided by human content moderators, with no explanation given as to why precisely certain content is removed. Through brute force testing, motivated actors can game the system to trick social media platforms into taking down legitimate content.

Unclear hate-speech guidelines for social media platforms, along with uneven enforcement of certain personalities, and a US-heavy approach compound the problems faced by international policy makers in creating effective legislation to combat online hate speech.

A multi-track policy framework, and a Moderation Standards Council (MSC), Dr. Tworek states, is Canada's best option in going forward against these challenges. This policy framework would provide an explicit definition of hate and harmful speech, giving legislators and law enforcement a clear picture of how this threat appears online. The MSC would be convened by the government and staffed by leaders of social media platforms, social scientists, and law enforcement members. It would address issues of fairness, accountability, transparency, and ethics as new challenges arise through online moderation.

Notable examples of attempts in combating online hate speech come chiefly from the EU. The Network Enforcement Act (NetzDG) in Germany created a framework for government to fine content providers (YouTube, Facebook, Twitter) up to 5 million euros for propagating fake news or distributing "agitating" content. A UK white paper has been published outlining the Duty to Care, emphasizing that safety in the workplace should apply to social media as well.

Dr. Tworek concludes that a meaningful course of action to combat online hate speech may be reached by empowering research through policy. Information gathering through AI and Big Data analytics will be useful but can present unknowing biases if not considered by researchers ahead of time. A database populated by information released by social media platforms, treated with the same sensitivity that healthcare information is, would be an effective tool in the creation of datasets to help identify problem areas in real time.

### **Question & Answer period**

When asked if Nazism/Identitarianism was being transformed by social media, Dr. Tworek offered that it was a possibility, although unsupported by current data. Common features and tropes of RWE remain grounded in identifying a “racially pure past,” and this has not changed, although research is underway as to the various mechanisms in how these groups gain strength and recruit.

Upon the discussion of specific targeting of members using hate speech in the broader context of VTSM, Dr. Tworek admitted it was a field requiring further study but offered the example of chat lobbies in videogames as an example. These games use closed-server, password protected channels so that specific groups can play together, specifically motivated actors appear to be using them for coordination.

In addressing if social media is changing how we participate in democracy, Dr. Tworek used the example of the 2016 US election, highlighting the process as a cry for help from people who thought democracy was failing them. Further, that rising inequality was contributing to a polarizing effect on the voting spectrum. The question was not if social media was changing how participation occurs, as this is now well established. Therefore, social media itself must be made democracyproof, so as to prevent motivated actors from projecting their will to such an extent that it will drown legitimate discourse.

On the subject of social media users choosing not to flag hateful content because they exist within an echo chamber, Dr. Tworek submitted that recent research is beginning to question the echo chamber hypothesis. She concedes that more research is required, but what is available shows asymmetry in echo chambers not only being a phenomenon of RWE, but of any extreme ideology. She also proposes that lay people may simply wish to know the perspective of the alt right, without having any specific affiliation for, or against them.

A question was raised that if wholesale takedowns continued to escalate in social media, law enforcement’s mission may be made proportionally more difficult. Dr. Tworek offered that this was possible, but that patterns existed where users who were banned on one platform would migrate to a smaller platform with poorer API, and thus more difficult to monitor. Facebook, as an example, uses the philosophy of remove, reduce, and inform to guide its content moderation protocols, with a heavy downplay on outright removal.

## Roundtable Discussion

The discussion period focused primarily on whether Canadian law should be adjusted to better address the threat of extremism and hate speech. One suggestion proposed that the current definition of hate speech should be enforceable online, with tangible punitive measures exacted on known belligerents. An emphasis on a grassroots-approach was highlighted, and the perspective that the exact letter of the law to combat hate speech matters less than the capacity for governing bodies to investigate and enforce the law itself.

It was commented that hate speech itself has begun to change, and that RWE are seeking legitimacy in a clearly false, albeit disguised narrative. RWE are switching from “white power” to “white empowerment”, and in so doing, trying to infiltrate existing conservative institutions where they can impose their rhetoric.

Within Canada, the Armed Forces are sensitive to the threat of seduction by a false narrative by RWE. Targets, vulnerable from PTSD and disenfranchised with government, may turn to RWE and offer their security access, weapons acquisition and handling, and combat-specific knowledge to these groups.

Discussion geared around how Canadian government policy may address hate speech and harmful speech moving forward and the unprecedented aspects of the threat. This revealed a large body of collective interest in creating individual engagement of the propagators of hate speech and challenging their victimhood narrative.

Disrupting the process of what transforms individuals into agitators for RWE, it was argued, is not a process that should be done on an institutional level. Deeper experience and engagement at the individual level is required, no matter how much of a logistical nightmare it may become. Public awareness geared towards facilitating this goal should be focused in schools, sports clubs, and other places where vulnerable persons may participate.

From this subject, the topic of victimhood as an RWE narrative was discussed. Playing the victim of the “white man beset at all sides by minorities” leads to a later justification of violence against agents of imagined oppression. This imagined threat and oppression has led to shootings and attacks, and as incorrect as the logic justifying them, the perception of threat is real and driving future violence, as people are rallying out of a fear to a threat that does not exist.

Evolving from this topic, the possibility of inoculation of current agitators and at-risk bystanders was discussed. Showing these at-risk people how the disinformation that they are being subjected to works, could be a powerful tool in not only preventing their own radicalization, but serving as an example to others. Examples of how this has been facilitated in the past were discussed, such as in Germany when members holding extreme left and right views were brought together and actually met, putting a face to the ideologies they had come to despise.

## **KEY POINTS OF DISCUSSION AND WEST COAST PERSPECTIVES**

### **Presentation**

- Social media is being used, more than ever before, as a force multiplier by motivated actors to disseminate hate/harmful speech online.
- Both individuals and large groups are vulnerable to targeting by motivated actors, thereby threatening the integrity of democratic discourse and exchange.
- Canadian efforts to combat hate speech are being stymied by private sector interests, which must be addressed if meaningful policy is to be made.

### **Roundtable Discussion**

- Attitudes regarding Canadian legislation governing hate speech reflect a strong desire for change in combating hate online.
- An approach that will consider the needs of law enforcement, government policy makers, private sector interests, and research initiatives is a strong candidate for implementation.
- Greater emphasis should be placed upon dispelling the disinformation tactics used by motivated actors in the pursuit of their goals, so that the actors may be removed from the cycle they are perpetuating.



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

© (CASIS-VANCOUVER, 2019)

Published by the Journal of Intelligence, Conflict and Warfare and Simon Fraser University

Available from: <https://jicw.org/>