

# DANGEROUS SPEECH PRE AND POST COVID-19: Countering Online Hatred While Protecting Freedom of Expression

## EVENT PRESENTED BY

Network for Responsible  
Public Policy

## EVENT CO-SPONSORS

The Academy of Political Science  
Women for Progress  
The Puffin Foundation, Ltd.

## SPEAKER

CATHY BUERGER is Director of Research at the Dangerous Speech Project (DSP), Research Affiliate of the University of Connecticut's Economic and Social Rights Research Group, Managing Editor of the *Journal of Human Rights*, and Editor for the Teaching Human Rights Database.

# DANGEROUS SPEECH PRE AND POST COVID-19

May 21, 2020

## ABOUT THE EVENT HOST



NETWORK FOR RESPONSIBLE PUBLIC POLICY is a non-partisan 501(c)(3) nonprofit organization that provides credible news and authoritative information on key public policy issues to foster informed citizenry and civic engagement. NFRPP is a trusted source that follows media literacy guidelines. It strives to motivate conversations, mitigate polarization, and combat misinformation. NFRPP's vision is to contribute to and foster a vibrant democracy through education and civic engagement. For more information, visit: [www.nfrpp.org](http://www.nfrpp.org).

## ABOUT EVENT CO-SPONSORS

**APS**

ACADEMY OF  
POLITICAL  
SCIENCE

THE ACADEMY OF POLITICAL SCIENCE, founded in 1880, promotes nonpartisan, scholarly analysis of political, social, and economic issues by sponsoring conferences and producing publications. Published continually since 1886, the Academy's journal, *Political Science Quarterly*, is edited for both specialists and informed readers with a keen interest in public and international affairs. For more information, visit: [www.psqonline.org](http://www.psqonline.org).

WOMEN FOR PROGRESS is a network of women whose mission is to raise awareness, educate and connect individuals to activism on a wide range of progressive issues including women's health, gun violence prevention, environmental protection, human rights and equality for all. For more information, visit: [www.womenforprogress.org](http://www.womenforprogress.org).

THE PUFFIN FOUNDATION, LTD. has sought to open the doors of artistic expression by providing grants to artists and art organizations who are often excluded from mainstream opportunities due to their race, gender, or social philosophy. For more information, visit: [www.puffinfoundation.org](http://www.puffinfoundation.org).

# Countering Dangerous Speech Online

CATHY BUERGER

DRAWING ON 18 MONTHS OF RESEARCH at the Dangerous Speech Project (DSP), Buerger discussed the challenges of responding to hatred online, as well as best practices.

## WHAT IS DANGEROUS SPEECH?

- Dangerous speech is any form of expression that increases the risk people will commit or condone violence against members of another group.
- Even in the largest scale incidents of intergroup violence, a small percentage of the population actually commits violence. However, a much larger percentage of the population allows it to happen—either by actively supporting violence or by just staying quiet.
- The concept of dangerous speech emerged from research into public speech that preceded several incidents of historical mass violence.
- Dangerous speech displays remarkably similar patterns across languages, cultures, countries, and historical periods.

## HALLMARKS OF DANGEROUS SPEECH

- *Dehumanization*. This is any kind of speech that refers to members of another group as less than human. Common keywords include vermin, rats, and insects.
- *Accusation in a mirror*. This term originates from a propaganda manual that was found after the Rwandan genocide. It is a tool used by speakers to convince people that violence is necessary. This puts people in the mindset of self-defense, and shifts violence from being a possibility, to being a necessity.
- *Invasion rhetoric*. The outgroup is portrayed as a contaminant and threat to the purity of the in-group.

## IDENTIFYING DANGEROUS SPEECH

- In addition to examining the message, it is imperative to examine the circumstances in which it reaches people.
  - *Speaker*. Does the speaker have any authority? Is the speaker charismatic? Is there a reason for people to fear the speaker?
  - *Audience*. Are there any factors that make people obedient to the speaker? Are they fearful?
  - *Social and historical context*. Is there a history of tension between the groups? Where is the power in the legal system?
  - *Medium*. Is the message spread in person? What language is used to spread the message?
- This is not a quantitative assessment—it is qualitative. Researchers gather as much information as possible, and make an educated guess about the likelihood that the messages will lead to people supporting violence.

## DANGEROUS SPEECH V. HATE SPEECH

- Dangerous speech usually conveys a sense of fear, but is not necessarily focused on hatred. Though, the two do overlap.
- There is hateful speech that can be dangerous. However, there is also plenty of hurtful and unkind speech that does not necessarily move someone towards violence. Buerger explained that the framework outlined above for identifying dangerous speech should be considered.

### DANGEROUS SPEECH AND COVID-19

- Anti-Chinese and anti-Asian speech has been on the rise as a result of the coronavirus.
- Beyond dangerous speech, there is also a rise in attacks, especially against Asians and Asian-Americans.
- Researchers have also been documenting increases in anti-Semitic, Islamophobic, and anti-immigrant speech. These groups are somehow being blamed for the pandemic, or for trying to profit from it.

### WHAT CAN WE DO ABOUT DANGEROUS SPEECH?

- The internet allows for one message to reach thousands or more within a minute.
- There is a general feeling amongst people that we just need to get as much of the dangerous and hateful speech off the internet and social media as quickly as possible.
- While taking this action is considered imperative, it can have ramifications in terms of freedom of expression. In countries with restrictive speech laws, removing dangerous speech is often applied unevenly. It is used as a way to silence political opposition and minority voices.
- The Dangerous Speech Project (DSP) is skeptical that the aforementioned action is the best way to proceed. As an alternative, people can engage in user-led responses by choosing to respond to dangerous speech when encountered.
- Counterspeech is any direct response to speech that seeks to undermine it. Facebook has touted this method as a way to “cure hatred” and combat extremism. However, there is not much research into counterspeech’s effectiveness.

### DOES COUNTERSPEECH WORK?

- It is very difficult to change the mind of someone who is posting hatred. Humans are very good at rejecting information that runs counter to what they like to believe.
- Counterspeech is effective at getting new counter-speakers to join the conversation.

- The number of counter-speakers involved in a particular conversation matters. As do the tone and personal characteristics of the counter-speaker. People are more likely to change their minds if the counter-speaker is like them in some way.
- There are many different methods, strategies, and goals to counterspeech efforts. The unifying factor is that counter-speakers are all working on the “condoning” group.
- Counterspeech might not have the power to convince someone who is willing to commit violence to not engage, but it can be really powerful at getting people to stand up and tell others to stop.

### ONLINE COUNTERSPEECH EFFORTS

#### *Dawud Walid*

- Muslim activist from Dearborn, Michigan, and author of the essay, “Fellow Humans are Not ‘Abeed’”.
- Abeed is a derogatory term in Arabic meaning “slaves” and is often used to refer to black Muslims. Walid was offended by the use of this word online and in his community. He knew the origins of the word, and felt people used it in a very casual way that did not represent the term’s history.
- Beyond writing this essay, Walid searched for the term “abeed” on Twitter, both in English and Arabic. He sent direct messages to Twitter users using the term, provided a link to his essay, and asked them to please read it. He received three types of responses: no response; doubling-down on usage of the word; or a positive reaction from individuals who learned the true meaning of the word.
- This represents a typical counterspeech interaction—person A encounters bad speech and says something to person B with the goal of educating. It is incredibly difficult to get someone to change their behavior. In Walid’s case, many of the people he was trying to reach were not intentionally hateful, but were unaware of the origins of the term.

*Mina Dennert*

- Iranian-born Swedish activist who in 2016 started the closed Facebook group, Jagärhär, meaning “I Am Here” in Swedish.
  - The members of the group post links to news articles seen on Facebook that have hateful comments on the posts. Members of the group then collectively go into the comments section and write counterspeech with fact-based comments and civil discourse that challenge dis- and mis-information. They also write supportive messages to the people being attacked by other commenters.
  - They tag their posts with #JagÄrHär and like each other’s comments. This allows their posts to be amplified based on Facebook’s algorithms. The group’s goal is to push their comments to the top and bury the hatred.
  - They hope to influence the larger conversation and make their posts more visible. Research shows that the first comments someone reads on a post have a contagious effect on the rest of the comments section.
  - The Jagärhär group has 75,000 members and has been replicated in 14 countries. Their goal is not to change people’s mind against posting hatred. Rather, the goal is to reach the movable middle, and inspire other counter-speakers.
  - One voice against hatred would get attacked. However, when a member is one of 100 people posting counterspeech, members feel braver in their efforts because of the group.
- The account now has over 370,000 followers. The goal is to educate others about how we are still in a nation with racism.
  - The account is also used to shame people who are posting racist things. Research shows that shaming is not a good way to get people to change their beliefs, but it occasionally gets them to change their behavior. It has yet to be studied whether this action drives people to other outlets where they won’t encounter counterspeech.

*@YesYoureRacist*

- Twitter handle started by Logan Smith after Barack Obama was elected.
- Smith was sick of people saying we are in a post-racial nation. He searched the phrase, “I’m not racist, but,” on Twitter, and was horrified at the number of people posting with this phrase.
- Smith started the handle, @YesYoureRacist, and began retweeting every post he saw with this phrase. The handle itself became counterspeech.