

How Online Harms Regulation Empower Speech and Engagement

Jonathon Penney

Osgoode Hall Law School // Citizen Lab, University of Toronto
Institute for Rebooting Social Media, Harvard University
Berkman Klein Center for Internet & Society, Harvard University

One of the [most common](#) and [forceful criticisms](#) of the Government of Canada's planned online harms legislation, including [early iterations](#) and [opinions](#) on [more recent](#) versions, is that such regulations will have a profound chilling effect on people's rights and freedoms, particularly online speech, sharing, and engagement.

These criticisms are not surprising. Such concerns [have often been raised](#) to criticize, oppose, or challenge laws and regulations aimed at addressing online hate, cyberharassment, disinformation, and other online harms. Not only that, social media platforms have been highly successful in weaponizing such claims to advance an anti-regulatory agenda by framing law and regulation [as a threat to users and innovation](#).

To be clear, chilling effects are real. I've documented and explored them in my own work, including the corrosive [chill of mass surveillance](#), [automated legal enforcement](#), or [online personal threats](#). But in the regulatory context, the evidence is far less clear. In fact, findings in recent empirical studies, including my own, demonstrate the contrary: regulations enacted to address online harms, like the forthcoming federal legislation, can actually have an *empowering* effect. They can encourage more speech and engagement online, especially by women and minorities. Any chill, by comparison, is negligible.

Chilling Effects and Our Permissive Legal Infrastructure

Concerns about chilling effects—that certain laws or regulations may “chill” or deter people from exercising their rights and freedoms—have long been central to debates about online content regulation and moderation. Often corporate service providers and platforms have employed them, with great success, to curtail regulatory efforts and promote a broadly permissive legal and regulatory environment.

Perhaps the best example is Section 230 of the United States' Communications Decency Act. This provides internet and social media platforms with near blanket legal immunity, shielding them from liability for user generated content and from lawsuits relating to how they moderate content. Section 230 is treated as a “sacred cow” by the U.S. technology industry. But it [has been controversial](#) as its broad legal protections now shelter powerful corporate platforms like Facebook, Google, and Twitter from legal accountability, while [providing little incentive](#) to address online abuse and other harms.

Chilling effect claims are a central reason for Section 230's broad scope. The U.S. Fourth Circuit Court of Appeal's famous 1997 decision in [Zeran v America Online](#), which provided Section 230 with its broad interpretation, was premised on concerns about chilling effects. AOL had argued that without blanket immunity, online service providers (OSPs) would chill and suppress online speech to avoid liability. The court bought that anti-regulatory framing entirely, [despite the fact](#) that Section 230's text, history, and original statutory intent did not justify *Zeran's* blanket immunity interpretation.

Despite this, courts since have been largely unwilling to disturb Zeran’s broad Section 230 reading. Lawmakers have likewise failed to enact Section 230 reforms, despite years of advocacy by critics, fostering a deeply permissive legal and regulatory infrastructure for platforms.

Given that most of the powerful and popular social media platforms today are American, the country’s permissive regulatory approach, underpinned by Section 230, has impacted Canada. We have no general intermediary liability statute to police platforms. Both courts and governments [have been very reticent to change that](#). When they have taken steps to do so, as the Trudeau government has experienced, [concerns, claims, and critiques about chilling effects resurface](#).

Despite their prevalence, there is little systematic study of such claims. Empirical research suggests that laws do not have the chilling effects that critics suggest. Furthermore, these criticisms also neglect *other* chilling effects—those caused online harm and abuse itself. Leading privacy and online abuse scholars like Danielle Citron [have extensively documented](#) how online harassment, bullying, and abuse have a profound chilling effect—a “totalizing and devastating impact”—that chills victims into silence, with disproportionate impact on women and minorities.

How Regulation Can Empower Speech

In one [article](#) published in 2019, Citron and I have explored how a cyberharassment law might impact what participants would be willing to say or do online. Using a study of nearly 1300 US-based adult internet users, we found that the cyberharassment law we tested would have a negligible chilling effect. Most participants indicated that the law would either have no impact or actually make them somewhat or much more likely to speak, share, and engage online. Not only that, we found the law would have an empowering effect—actually *encouraging* these activities, particularly for women.

We explained the findings using expressive law theory—a growing body of behavioural research that focuses on the [expressive function of law](#)—how it can shape behavioural norms by changing the social meaning of behaviour. When a law is passed, it provides a powerful symbolic or “informational” signal as to societal consensus or wider popular attitudes about social behavior, meaning how people should act and what behaviour is approved and disapproved. The law also provides information about the relative “risk” of certain behaviour. New online regulations raise the risk of abusing and reduce the risk of speaking and sharing online, especially for those most often victimized by online abuse. Over time, people internalize the attitudes and norms expressed by the law, altering broader behavioural norms.

Given that women are disproportionately targeted by online harassment and abuse, our findings made sense in terms of expressive law theory. A cyberharassment law designed to deter online harassment and abuse suffered by women had a positive expressive effect on women’s speech, sharing, and engagement online.

Our more forthcoming recent experimental research explored the impact of legal and platform measures aimed at protecting intimate privacy from abuse and invasion. We similarly found negligible evidence of any chilling effects. We *did* find that these privacy protective measures promoted trust, which is critical to fostering greater intimate sharing and expression, both online and off, especially among women and minority groups more likely to be victims of intimate privacy threats and similar abuse.

Our findings are consistent with other recent studies like computational social scientist Nathan Matias, [who found](#) that rules concerning online abuse in online communities helped curb online harassment and encouraged wider group participation.

Implications for Online Harms Law and Beyond

So, the critics are wrong. There is little evidence to support claims that online harms regulations—like Canada’s planned online harms laws—would chill speech and engagement online.

Our findings showed the very opposite: online harms legislation, if carefully tailored and communicated effectively, can *support* and *encourage* a wider diversity of speech and engagement, especially for those most often silenced due to being targeted for abuse—women and minorities. This is the empowering effect of online law and regulation.