FOR WHOM THE BELL TROLLS

RYAN CALO*

Hate Crimes in Cyberspace is among the most important academic books about the Internet in recent memory.[1] It is creative, rigorous, and almost implausibly well written.[2] The characterization of the underlying issue of online hate crimes is quintessential Danielle Keats Citron: stories to move, data to prove. You couldn't read this book and think the problems Citron identifies should be ignored. And her solution—better, and better enforced, laws—is clear, pragmatic, and entirely plausible.

My comments amount to a simple observation: not all trolls are alike. Citron pays appropriately significant attention to the victims of online hate, treating these individuals in all their depth and variety. Her portrait of the perpetrator is thin by comparison. Chapter Two (pages 56-72) discusses the various forces that foster and exacerbate cruelty online. But the person who engages in the kind of hate speech Citron is concerned with—the sort that merits criminal prosecution—does so with a vehemence, cruelty, and diligence that would frankly be impressive in other contexts.

These Super Trolls really exist, and in regrettable numbers. But so do bored teenagers looking to push limits. So do mentally ill people. Increasingly, so do speakers who are not really people at all.

To the victim, it likely does not matter what the source of the hate speech is. A death threat is just as scary if it's some suburban, basement-dwelling kid—himself the subject and product of bullying. It's even scary if there is no human source at all, as a recent case in the Netherlands illustrates. Police visited a man because of something they read on a Twitter feed registered to him that sounded like a specific threat of violence to an event in Amsterdam. It turned out that the

* Assistant Professor at the University of Washington School of Law; Faculty Co-Director of the University of Washington Tech Policy Lab.

[1] DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE (2014). Citron is one of my favorite people in the multiverse. Just because I'm biased, however, doesn't mean I'm wrong. Think of your paranoid friends who said the government was monitoring all of our communications—you owe your friend an apology.

[2] Look, for instance, at the last paragraph beginning on page 197: the way Citron dismantles a century-old argument merely through her subtle and fair characterizations of the supposed claims of hate speech. *Id.* at 197-98. I suffered a physical pang of jealousy reading that paragraph and many others.

tweet was an example of emergent behavior by a software program.[3] Neither the programmer nor the operator intended the bot to threaten anyone.[4]

When it comes to the Super Troll, we are in fact talking about a civil rights issue, and the appropriate response will involve law. But is the rhetorical and policy calculus the same when the perpetrator is young or mentally ill, let alone inhuman? Kids may not be aware of the law or appreciate its "expressive value."[5] A jury may be reticent to penalize a scared and apparently repentant perpetrator. A bot has no body to kick or soul to damn.[6]

Now, exactly who is generating what hate speech on the Internet remains an empirical question. But for some sources of hate speech—whose attacks, again, land just the same way with the victim—different solutions will be needed. I'm thinking here of how Riot Games uses machine learning to detect and penalize abusive language within the massively multiplayer online role-playing game League of Legends, apparently to genuine and lasting effect.[7] I'm thinking of interventions in schools. I'm thinking of means, such as bot registries, to check whether the originator of a comment is actually a person.

At the same time, we should resist the intuition that these "code" or disclosure-based solutions will somehow clean up the Internet on their own.[8] Just as the Minor Troll might not respond to the law, the Super Troll may respond to little else. These folks are dedicated in their hate; they will ignore norm entrepreneurship and find ways to end run technical controls. For this population, I agree with Citron: throw the book at them, edge first.

---

[3] On emergent behavior and law, see Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 538-45 (2015).

[4] *See* Kashmir Hill, *After Twitter Bot Makes Death Threat, Its Owner Gets Questioned by Police*, FUSION (Feb. 11, 2015), http://fusion.net/story/47353/twitter-bot-death-threat/ (citing to comments made by bot developer and owner of the relevant Twitter account). Something like sixty percent of Internet traffic is bots. Leo Kelion, *Bots Now 'Account for 61% of Web Traffic,'* BBC NEWS (Dec. 12, 2013), http://www.bbc.com/news/technology-25346235.

[5] *See* Danielle Keats Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 MICH. L. REV. 373 (2009).

[6] *Cf.* Peter M. Asaro, *A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, *in* ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTS (Patrick Lin, Keith Abney, and George Bekey eds. 2011).

[7] *See* Simon Parkin, *A Video-Game Algorithm to Solve Online Abuse*, MIT TECH. REV. (Sep. 14, 2015), http://www.technologyreview.com/news/541151/a-video-game-algorithm-to-solve-online-abuse/.

[8] The title and several passages of *A Video-Game Algorithm to Solve Online Abuse* suggest that the author or interviewees believe at least some categories of "toxic behavior . . . can be addressed through a combination of engineering, experimentation, and community engagement." Id.